

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <BEV>
A:Cross-references: EMBL:AL153994; GSPDB:GN00063; ATSP-F17114.60
A:Experimental source: cultivar Columbia; BAC clone F17114
C:Genetics:
A:Gene: ATSP-F17114.60
A:Map position: 5

Query Match 37.7% Score 397.5; DB 2; Length 224;
Best Local Similarity 46.6% Pred. No. 5,1e-27;
Matches 88; Conservative 28; Mismatches 46; Indels 27; Gaps 3;

QY 22 SNHSFKRSMPEQPHLLMDWKNKANDLTQEHAAFLNDPRLHMLD----- 65
DB 2 NNNKMPNLSLFQNTYNNNNIINSSNNNNKNDHGHNNNDPJGAMDOYTOHIFNPFSSR 61
QY 67 -P-ETLIHDEDEYD-----EDMDAKEMQYIAVMPVDIPATVPKPNR 114
DB 62 FPPILSSILTTTLTSGDDDDDEEPLDELGAKEMMYKIAAQSVYDIDPATYKPKRR 121
QY 115 NVRISDDPQIVVARRRERISEKIRIKRIVGCAKMDTASMLDEAIRTKKQVRIL 174
DB 122 NVRISDDPQIVVARRRERISEKIRILQIVGCTKMDTASMLDEAIRTKKQVRIL 181
QY 175 QPHSQICAP 183
DB 182 NNTGTTPP 190

RESULT 3
143579
hypothetical protein F11C1.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
R:Accession: T45579
R:Barques, M.; Collado, M.C.; Navarro, P.; Tercil, J.; Perez-Alonso, M.; Mewes, H.W.; May
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23007
A:Accession: T45579
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <BAR>
A:Cross-references: EMBL:AL132976
A:Experimental source: cultivar Columbia; BAC clone F11C1
C:Genetics:
A:Map position: 3
A:Note: F11C1.170

Query Match 34.1% Score 360; DB 2; Length 231;
Best Local Similarity 45.8% Pred. No. 1.5e-23;
Matches 82; Conservative 26; Mismatches 37; Indels 34; Gaps 6;

QY 22 SNHSFKRSMPEQPHLLMDWKNKANDLTQEHAAFLNDPRLHMLDPEPTLIHDEDEE 80
DB 23 SNSNP-----NPNPHNIM-LSES-----THPFFNPTHSLPDP---QTPRHQPOLN 68
QY 81 Y-----DEMDAKEMQYIAVMPVDIPATVPKPNRNVYISDD 121
DB 69 FRVAPSSSLPEKRGCGSDNANMAAMREMIFRIAYMQPIHIDPESVPPKPKKNRISKD 128
QY 122 POTVARRRERISEKIRIKRIVGCAKMDTASMLDEAIRTKKQVRILQPHSQI 180
DB 129 POSVAARRRERISERIRIQLVPGCTKMDTASMLDEAIRTKKQVSLSHNAV 187

RESULT 4
T48153
hypothetical protein T1008.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48153
R:Barques, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Maye
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24486
A:Accession: T48153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-912 <BEV>
A:Cross-references: EMBL:AL161746
A:Experimental source: cultivar Columbia; BAC clone T1008
C:Genetics:
A:Map position: 5
A:Introns: 166/1; 260/3; 551/2; 750/3; 807/3
A:Note: T1008.20

Query Match 18.1% Score 191; DB 2; Length 912;
Best Local Similarity 61.3% Pred. No. 2e-08;
Matches 38; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 110 KPNRRNRISDDPQIVVARRRERISEKIRIKRIVGCAKMDTASMLDEAIRTKKQVR 169
DB 35 KRKSTSTLSDPOSVAARRRERISOREKILQSVPGCAKMDVSMDEAIRTKKQVRIL 94
QY 170 QY 171
DB 95 QI 96

RESULT 5
C56690
unknown protein F28G11.9 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Accession: C96690
R:Barques, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Maye
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24486
A:Accession: C96690
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <STO>
A:Cross-references: GB:AE005173; NID:q11054535; PIDN:AA627834.1; GSPDB:GN00141
C:Genetics:
A:Gene: F28G11.9
A:Map position: 1

Query Match 16.9% Score 178; DB 2; Length 298;
Best Local Similarity 25.9% Pred. No. 6.6e-08;
Matches 55; Conservative 30; Mismatches 69; Indels 58; Gaps 6;

QY 17 SSKRSNHSFKRSM---EPQPH-LMDWKNKANDLTQEHAAFLNDPRLHMLDPEPTLIHDEDEE 56
DB 50 STNNSDQHQCGVFYPSGCDHNSLMDPNCSSFLFDHHSFPRAISCGSGGCGF 109
QY 57 -----LNDPRLHMLDPEPTLIHDEDEEYEDMDA 87
DB 110 SFLGNNNSYGTFTNHNHNDHDIISPSTETPGQCKDMLYSDSTVTTTSRNSLSPSA 169
QY 88 MKEMQYIAVMP-----VDIPATVPKPNRNVYISDDPQIVVARRRERISEKIRILK 142
DB 170 GKRRSHTEGSPQSKLSSGYTKTKPKPTTS-----PKDPSIAAKNRERISERIKILO 225
QY 143 RIVPGCAKMDTASMLDEAIRTKKQVRIL 174

gene
complement(23534..25288)
/gene="F6N15.15"
complement(23534..23629,23711..23845,23934..23999,24045..24092,24457..24490,24753..24823,25132..25288)
/gene="F6N15.15"
/note="contains similarity to Lotus japonicus RING-finger protein (GB:236750)"
/codon_start=1
/evidence="not_experimental"
/protein_id="AAC19316.1"
/db_xref="GI:3193334"
/translation="MSLSENEELFDINLEIPEKTVSDGDDEFDFLNKIPREFTVNSSEDFADHLHRNQATRRKKLRKPRKPRHLEKVISSEFLVRSASFLELDLNMGLRMDTPYOELRMDTDHMTLEOLICNNMGYENGSGKASNDKLRNKRKSHFSIADKICICODPGFQRAGYGRKINCGHNFHVNVKMITTK"

gene
26808..27437
/gene="F6N15.9"
26808..27437
/gene="F6N15.9"
/note="contains similarity to pectinesterases"
/codon_start=1
/evidence="not_experimental"
/protein_id="AAC19307.1"
/db_xref="GI:3193325"
/translation="MAANNKLFVLSLPIIFSAATATSKYDTRKAYHSCRTILYKLCVRSRVSRAVONPRDLARFLKASLIRAVTAFLKEVKNLETLRQYVASVHDCLOLRSDYSLAIEIDRVSROGSGDLHWHINLQCTSLTIDAEICVSOFPGRKMSKLAITKGRVKEETSNALAFIEHVAARYRARRP"
27662..27734
/note="codon recognized: AAG; Lys"
/product="LRNA-OTHER"
28096..30456

CDS
59.6%; Score 3842; DB 8; Length 95643;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match
8 ccaactactaccggttctgtaaccccttaagcttgaagactactagaataatgaacaat 67
Db 39816 CCACCTACTACTCCGGTGTGACCCCTTAAGCTTTGSAAGACTACTAATAATAGCAAAAT 39757

Qy 68 accataatgccataccatcccttcttcttctgtaacatgaacatcttaatttgaataa 127
Db 39756 ACCATATGTCCATATCCATCCCTTTCTTTGTTGAACGACATTTCAATTTGTAAAA 39697

Qy 128 gaaaaaacctatgtaataatcacctgaagcaaaaaaataatcatcatataatattt 187
Db 39696 GAAAAAACCTATGTTAATATCCCTAGCAAAAAAANATATCTCATATTTAAATTTT 39637

Qy 188 tatataagaattacacatccctcgtctgaagcttaactccaattgnaagtgctgtaata 247
Db 39636 TATTTATAGATTATACATTCCTCGCTTGTAAGAGTTACTCCAAATGCCAAGTTGTATTA 39577

Qy 248 actaaataaaagcaagaatgaagcaactataatgaattgtaattgtaataatgaata 307
Db 39576 ACTATATAAAGACGAAGCTTAGAAGCTTATATTAATGATGTTGCAATAGTACTGCTA 39517

Qy 308 tatctgtgatgaataatacaagatagaacattatgcatgaacggggatcttgccttg 367
Db 39516 TATTTGTGATGAATATAAAGATGTGACATTAATGATGAACGGGCTATTTTGTCTTG 39457

Qy 368 aactcatlaaagcaatgtaagaagaatgtaagctcatcttgaataatttcaatctct 427
Db 39456 AACCTCATTAAGCAATGTGAAAGAGATGTGAGCTCTCATTTTGAATAATTTATCTTCT 39397

Qy 428 agctt-gtgcatttcaaacctatgaagaaagcaacatatagaataattcattgtaagaa 487
Db 39396 AGCTTTGTGATTTTAAACTTAATGAATGAAGCAACATATAGAAATTCATGTGTCACAA 39337

Qy 488 cgaacttagacggtatcctaataatagaacgaatcaatagtaataatactatatataat 547
Db 39336 CCACCTTTAGCGGTATCTTATTAATAGCCCATTAATATAGTAATATATATATATAT 39277

Qy 548 taagtgtatataagtttacttataccacttgaataattaaacaatgagcaactctaa 607
Db 39276 TAGTGTGATATATTAAGTTACTTATCCACTTGAGAAATTAACCAATGGCAATACCTTA 39217

Qy 608 tctgaaagaagccgtcccaactcgtgtataatgaattagggggaagatcctgttaaat 667
Db 39216 TSTCCAAAGAACCGTCCCACTTCGTGTATGATGATATGGGGAAGATCCGTTAAT 39157

Qy 668 cgtcaaataaacaacttaagaataatgacacacaaataatcataagaagacgttg 727
Db 39156 CGTCAAATTAACCACTTAAGAACTAGAAATTAATGACACCAAAATCTATAAGAGAACGTG 39097

Qy 728 aagaagctatcatcgatccagctatcttccctgaagaatgaataatgaagccgttg 787
Db 39096 AAGAACTATTATCGATCCAGCTCATATTTCTTACGTAGATCAAAATCAAGGCCGTG 39037

Qy 788 aaaggtctgtaagaataatgtaagaacacgtgaggtctagaagaagacaagaatag 847
Db 39036 AAAGCGCTTGTAGAAATATGCGAAGAACCGTGGGTTAGAGAAAGCAAGAAATAG 38977

Qy 848 aagaacaatgattgaatctgctatcttctgtaagaagttgtcaaaagagagagag 907
Db 38976 AAGAACTATGATCTTAAATTCCTATTTTGTGTATAGAGTTGTCAAAAGAGAGACAG 38917

Qy 908 aagaagaataatagttcaaaaataatgaagcaactaaatgaagacatgtgtgaactat 967
Db 38916 AGAAAGAAATAGCTCAAAATTAATGACACCTAAAAATGAGACATGTGTGATTAAT 38857

Qy 968 laaagaagcaactatccttccctatagcaatgatctccaaacaagtgcaagcctct 1027
Db 38856 TACAAAGACGACTATGCTTCTTATGCAATGATATCCAAACCAAGTCCAAAGCTCT 38797

Qy 1028 tcttgcctaatctcglaaagctctcctccttctccttcttcttcttgaagaaacctaga 1087
Db 38796 TTTTGTCCCTATTTTTCGTAAGATCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 38737

Qy 1088 tttaacccctgtcttctgaatcttcttcttgaagaacatgatcttgacacacactat 1147
Db 38736 TTTAATCCCTGTCTTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 38677

Qy 1148 tcttcatcttctgtgctatagaagatttgccttaataatgcttcttcttcttcttctcgt 1207
Db 38676 TCTTCTATCTTTTGTGCTATAGATTTTGTCTTATATGTCTTCTTCTTCTTCTTCT 38617

Qy 1208 acgtacgtatcgaatlaaattggtlaaacaaggttataataactagaacaagaatgag 1267
Db 38616 ACGTACGTATACGAATTTAAATGCTTATTAACCAAGTTTATATTAACATGCAAAATGAG 38557

Qy 1268 ccatgaaatctgttagcgaanaagtagaataatlatgaagcttcaaacggaatatata 1327
Db 38556 CCATGAATAATTTGTTAGCGAANAAGGTAGAAATATATGATGCTTTTAAAGCAATATATA 38497

Qy 1328 taacttgcgaacactgaacttataatcgaatgaataatcttcttcaatctcct 1387
Db 38496 TAATTTGCTGCAAACTTACCTTATATCATGATCTATATATATTTTCTTTATTTCT 38437

Qy 1388 tgcgaatlaatcaatgacaggaatcttgcgaagtatgctgcgaatcttccatccac 1447
Db 38436 TGCCAAATTAATACATGACAGCGATTTTGTGCAAGTTATGTGCAATCTTCCATTCAC 38377

Qy 1448 acaacacacttaacttaagaactctgaagaatattttaaataagaacacttatacga 1507
Db 38376 ACAACACTAAACTTAATTAAGACTCTGGAATAATTTTAAAGACAACTTATTCGAAA 38317

Qy 1508 aacttagtatagaanaacactccagaatlaaacaatgaagctataataattgaataa 1567
Db 38316 AAATTTACTTATGAAAACATTTCCAGATTAATTAACATGAGCTATATTAATTAAGATAAAT 38257

Qy 1568 gaagtaatatgtatgataataatacatatcgtctgcgtaaaaaaaacatatctg 1627

Db	38256	GAACTATATGTGATAGTATGTAATACATATCTGATTTGCCGTAAAAAAACATATCTG	1687
Ox	1628	atlaaatlgttcacggaagcccaatgctcaatagatcatcaagtttttatttccaaat	1587
Db	38196	ATTAATATTGTCATGACGCCCATGCACTATGATGATCATGCTTTTATTTCACAT	38137
Ox	1688	aactcaatataatccaataaaatagtttgtcagatlaaatttttttggtcagctt	1747
Db	38136	AACCAATATATATCTCAAAAAAATAGTTTGCAGTTAAATTTTTTGGTGCAGCTT	38077
Ox	1748	ttcccaaccacaaactagtttggaaagttctctcttaattttctttcttgaatt	1807
Db	38076	TCTCCAACTCCTACAAACATGTTTGGAAAGTCTCTCTTTATTTTCTTTCTATTT	38017
Ox	1808	cttaatgtttttatcttaagaaatttaagcggatgttgtagtcgtttctctttct	1867
Db	38016	CTTATGTTTTTTTATTATGGAATTTTAAAGCCGATGTTTGGTCGTTCTCTCTTTCT	37957
Ox	1868	tgatttctaaagttacattctgttaaacatcatcccccacatlaagctcaatcatatg	1927
Db	37956	TGTTTTCTAAAGTACTTTTGTAAACTATCTCTCCCATTTAAACATCAATCAATAG	37897
Ox	1928	ttacctttaatatatgtctagtcgataaaaaaagaaaaaacatcgcggtagttcta	1987
Db	37896	TTATCTTTTAAATATGCTACTTATATAAAAAATGAAAAAATACGCGTAGTTCTTA	37833
Ox	1988	ctaagtctgttgraaaaaactcgaattatgaattctatcaattcttttgatgataat	2047
Db	37836	CTAATGTTTGTGAAAAAACTGATATTATGATCTAATCAATTTCTTTATCGTATAT	37777
Ox	2048	gtgggttaaatcttagtaatttttcaataaaagaaactgaattgttagtataatgg	2107
Db	37776	GTGGGTAAATTTAGTAATTTTTCATATAAATAGAACTGTGATGATATATGGG	37717
Ox	2108	gaatcagatataatagcttggtgtaacctactcttggaaactcttgaagattaattt	2167
Db	37716	GAACTACTATATTAGCTTGCGTAACATACTCTCGAAATCTTGAAGATTATCAATTT	37657
Ox	2168	gcaaaactataactttagtcctcgaaaaaatacagacagcagacacataaagcag	2227
Db	37656	GCAAAATATATATTNTGTCGCCAAAAATACACACGCGGACACGACATATTAACCG	37597
Ox	2228	gtttgaactcttggaaaactttgtbatacacaactataaatactaaagtcttgctgg	2287
Db	37596	GTGTGAATCTTGCAAAATTTTGTATACATAACCTATATATATCTAATGTCTCGTGG	37537
Ox	2288	ttcaaaagccttttcaaaagttccgctttttaaatcaagacactttacatagaaata	2347
Db	37536	TTCAAAAGCCTTTTCAAAAGTTCCATTTTAAATCAAGCACATTTTACATGGAATA	37477
Ox	2348	agttgagtcataaaaaaataatgatttattcttggaaagtttttttgaataaaagcaa	2407
Db	37476	AGTTGAGTCTATAAAAATATATGTTTATTTTGAAGTTTTTTTGTATTAACCGCAC	37417
Ox	2408	tattaagaagttagtttttttctactaccaaatatgaatttaaaacctatgacaaca	2467
Db	37416	TATTAACAACTTNGTTTTTTTTCACATCCAAATATCAATTATTTTAAACCATGCAACA	37357
Ox	2468	ttcatlaaaacaaataacatlaaagaatatlaaataatcacaataataataacttaaa	2527
Db	37356	TTCTATAAACAAATACATATTAAAGATATATAAATACACAAAATTTTAAATACACTTAAA	37297
Ox	2528	tttaacatatataattcaacaacacttaattatigaaaacagaaagaaagftaaaat	2587
Db	37296	TTTACATATATATTTTCAAAACATCTAATTATTTGAAACACAAAAGGAAAGGTAAAAAT	37237
Ox	2588	atcataaaatgagacatatatccataaaaaaaaatagcagcatgaagtaaatataa	2647
Db	37236	ATCTATAAATAGACACTATATCTCTTAAAAAAAATATAGGCATATAGTAAATATATA	37177
Ox	2648	gagacatgcatgaaacattcggctcaattaaatcgatcaagaatatataatcagtaaat	2707

Dh	37176	GAGCAGTCATGTAAGACCTTCGGTTAAITTAATGACGTCAAGATATATATACGTAAATATAC	37112
Qy	2708	atatgltgataattcttcgtaaaaaagaatataatatatctggaataataagaagatgtaaatg	2767
Dh	37116	ATATGTATATATTCTTGCAAAAAAGAAATATATATTATGAGAAATAGAAAAAGATCAAAATG	37052
Qy	2758	gaataatgataataaaaaagaagaagatgtagcgaatccttgatgctcgtcccaaaacaga	2827
Dh	37056	GAAATATGATATATAAAAAGAAAGAGAGTCCGACCTCTTGCTGCTGCTCAAAAGAGAG	36927
Qy	2828	tcgaaccagagcccacaaaanaacatgatgtagagagccttgagctcaaatctctcatatgat	2887
Dh	35936	TCCAACACACACCCCAAAAAGAACATGATGGAGGCTTCAGGCTCAACATCTCTCTCAAGAT	36933
Qy	2888	tgaacacaaagctaatgatactcttcacacacaagacacagcagctcttcataatgatacctac	2947
Dh	36936	TGGAAACAAAGCTATATGATCTTCTTCACACAAAGAACAGCAGCTTTCTCAATGATCTCTAC	36877
Qy	2948	catctcatgttaaatccactcccgaaaccccaattgcctcttgagcagaagcagaagatcac	3007
Dh	36876	CATCTCATGTTAGTACCTCCCTCCGAAACCCATTACTCTTGAGAGGAGACAGAAAGATCAC	36811
Qy	3008	gataaagaacatgagatgagatgaaagagatgcaatgataatgatacgcgtatgacgcctga	3067
Dh	36816	GATGAAGACATGATGCTGAGATGAAGAGAGATGACGTACATGATGCCGCTCATCACACCCGTA	36757
Qy	3068	gaactgaacccctgcgaacaggtcccttaagccgaacccgctgaacgaagataaagcaagat	3127
Dh	36736	GACATTCACCCCTGGCACCGCTCCCTTAAGCCGAACCCGCTTAACGTAAAGATTAAGCGAGAT	36697
Qy	3128	cctcagaagatggtgtgtgtctcgtcgtcgtcgaagaagatcacagagaagatccgaattctc	3187
Dh	36636	CCTCAGAGCGTGTCTGCTCTCGGCTCCGGAAGAGATCACAGGAACATCCGAAATCTCTC	36633
Qy	3188	aagagatcgltgectgtagtgcgaagaatggaacagcttccatgctcgcgaagaaagacata	3247
Dh	36636	AAGAGATCGTGGCTGCTGCTGCTGCGAAGATGAGACACAGCTTCACAGCTTCAGAGAACCTATA	36577
Qy	3248	cgttacaccaaagttcttgaagaagcagatgagaattctctagcctcaactccaaattgga	3307
Dh	36576	CGTTACACCAAGTTCTTGAACACGGCAGGAGAGATTTCTACAGCTCTACCTCTCAGATTGCA	36511
Qy	3308	gctccatagtgtaaacccctcttaaccttcttataaccacaactcccaacccctgacgac	3367
Dh	36516	GCTCCTATGGCTTAACCCCTCTTACTTGTATATATACCAACTCCCAACCCCTGATGACT	36455
Qy	3368	acaagaagctcgtgtgctagaaattggtgtcatccctccaaaccttccatgctgata	3427
Dh	36456	ACACAGAGCTCCCTACTACTAGACATTTGGTGTCAATCCCTCAACCTTTTCACTGTTGATA	36397
Qy	3428	tattatatagatgataaagaatcgtgaccaaagtcttatggctgtttaatatata	3487
Dh	36396	TATTATATATAGATGCCATAAAGTTGATGCCAAGATGTATGGGTGTTTAATATATATA	36333
Qy	3488	ttctaagatalatagatgatacaatgltgtaccaaagttccttatacctgatalcatatcat	3547
Dh	36336	TTCTAAAGATATATGATGTACCAATTTGTGTACCAAGTTCTTTATCTTATATATCAATATGCAT	36277
Qy	3548	aaataatctgtgataaagaagaagatgatgactgtgaacacaaaagaagaatattgat	3607
Dh	36276	AAATATATTGATGAAATAAAAAGAAAGATATTGATTGTAACCAAAAAAAGAAAGATTTGATT	36212
Qy	3608	gtaataagagttgatacatctcgtatgaagaacttgagccctgcacaaataatcttcatat	3667
Dh	36216	GTATATTAGGTTGATGATCATCTCTATGCAACCTTGCCCTGCAAAATTAATTTCGATAT	36155
Qy	3668	atataatatataggagaatataatacaaatctttaaattgacttaaatattgat	3727
Dh	36156	ATATATATATATATGCAATATATATACCAATCTTTTAAATTGACTATATATTTGTAAT	36097
Qy	3728	caattaccatcaatcagatgagttaagatataatagatatgacattgacaaaagaacacaca	3787
Dh	36096	CAATTATCTAAATCTGATGAGTATAGTATATATATGAGTTTGCAAAAAAGCAAAACACCA	36033

C:Accession: B66432
 R:Phooloosis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, C.; Alonso,
 Chin, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Jensen, N.F.; Hynes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.D.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, C.S.; Maitl, K.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: B66432
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-264 <STD>
 C:Cross-references: GB:AE005172; NID:94587523; PIDN:AA02574.1; GSPDB:GN0141
 C:Genetics:
 A:Gene: T518.12
 A:Map position: 1

Query Match 13.0% Score 137; DB 2; Length 264;
 Best Local Similarity 24.7% Pred. No. 0.00018;
 Matches 45; Conservative 34; Mismatches 63; Indels 40; Gaps 8;
 QY 26 PKRSMEPOPRLHLMNKKANDL---LTOEHAFLINDPHL---MLDPPPTLHDEP 78
 DB 45 PSQELFIPSPKTRV--FNESELDSPFTHQKLISSRHHVSHOFSFSPSNLDP-- 100
 QY 79 EYEDMDAMKEMO-----YMAVQPVLDIPATVPKRRNRVRISDDPGYVARRR 131
 DB 101 -SYTEASNIKRFQAPDFSSIFKYGWTEQD-----TKREL---SAOSIARRRR 147
 QY 122 ERISKIRILKRIYVPGAKMDTASMLDEAIRYTKFLKROYRILO-----PSQIGA 182
 DB 148 RRIKTKLOELKLIPOGSKINTAEEMFAAKYKFLQAOIEILQKOTMOTLDSKYGR 207
 QY 183 PM 184
 DB 208 EM 209

RESULT 14
 T49982
 hypothetical protein F12B17.80 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 *sequence_revision 02-Jun-2000 *text_change 02-Jun-2000
 C:Accession: T49982
 R:Bevan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the protein sequence database, April 2000
 A:Reference number: 225026
 A:Accession: T49982
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <BEV>
 A:Cross-references: EMBL:AL353955; GSPDB:GN0063; AUST:FL2B17.80
 C:Experimental source: Cultivar Columbia; BAC clone F12B17
 C:Genetics:
 A:Gene: ATSP:F12B17.80
 A:Map position: 5
 A:Introns: 179/3; 231/3; 287/3

Query Match 12.9% Score 136.5; DB 2; Length 315;
 Best Local Similarity 28.7% Pred. No. 0.00024;
 Matches 50; Conservative 23; Mismatches 70; Indels 31; Gaps 6;
 QY 17 SSKSKSN-----HSPKRSMEPOPRLHLMNKKANDL---LTOEHAFLINDPHLMDP 67
 DB 55 SSSSLMNIJSQPLHQP-----PQSSPLVD---SPLSSAFDYPFLIEDITHSYSP 105
 QY 68 PPELILHDEDEYEDMDAMKEMOYMAVQPVLDIPATVPKRRNRVRISDDP--QTV 125

DB 106 PP-LILPASDENNNISPLMEESKSFISGE-----TNKRKSKKLEGGPSKML 154
 QY 126 VARRRRRISKIRILKRIYVPGAKMDTASMLDEAIRYTKFLKROYRILOPSQ 179
 DB 155 AERRRRKRLNRLSLRSIVPKITKMDRTSLDADIDYKKEILLDKINKLOEDEQ 208

RESULT 15
 T00464
 hypothetical protein At2g34820 (imported) - Arabidopsis thaliana
 N:Alternate names: hypothetical protein F1913.5
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 *sequence_revision 12-Feb-1999 *text_change 23-Mar-2001
 C:Accession: T00464; C84761
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.
 submitted to the EMBL data library, April 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.
 A:Reference number: 214160
 A:Accession: T00464
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-266 <ROU>
 A:Cross-references: EMBL:AC004238; NID:93033373; PID:93033378
 A:Experimental source: Cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii,
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, I.; Tall
 euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ven
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: C84761
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-266 <STO>
 A:Cross-references: GB:AE002093; NID:93033378; PIDN:AA012822.1; GSPDB:GN00135
 C:Genetics:
 A:Gene: At2g34820; F1913.5
 A:Map position: 2
 A:Introns: 202/3

Query Match 12.9% Score 136; DB 2; Length 266;
 Best Local Similarity 50.9% Pred. No. 0.00022;
 Matches 27; Conservative 10; Mismatches 16; Indels 0; Gaps 0;
 QY 123 QTVARRRRRIRISKIRILKRIYVPGAKMDTASMLDEAIRYTKFLKROYRILO 175
 DB 143 OSIAARRRRIRIAEKTHLKGKLNLTAEWFOAAKYVFLQSOVGLQ 195

Search completed: March 29, 2002, 14:52:54
 Job time: 2617 sec

Db 226 ELVFNCTKYD:VTMLEKAIKSYKFLQVOVKYL 257

RESULT 6

B86402

protein T22C5.19 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86402

R:Phylogeny: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kall, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Hutzar, L.

N:Ref 408, 815-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.; Schwartz, J.R.; Shinn, P.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Southwick, A.M.; Sun, H.; Tallon,

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: B86402

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-279 <STO>

A:Cross-references: GB:AF005172; NID:96693022; PIDN:AAF24948.1; GSPDB:GN00141

C:Genetics:

A:Gene: T22C5.19

A:Map position: 1

Query Match

Best Local Similarity 16.68; Score 175; DB 2; Length 279;

Matches 60; Conservative 44; Mismatches 63; Indels 104; Gaps 10;

QY 13

13CVSSKSRSGNSPKRSM-----MEQPHLLMDNKKANDLTOEHAFL----- 57

19 DOSSSSK-----EERPRDELICSLYNGHLMHQH-----NNVTSIHHAFLLPDMF 67

58 -----NDPHLLMDPPET-----LIHLDEDEVEDMDAMKMQYM 94

68 PFGAMPQGNLPAFLMDSDQSHLQ-----ETSLKRLDVENLCKTNSCDVTROELAK 122

95 IAWMPVDIPATVPKPN-----RRNAISDDPOTVYA 127

123 SKKKQVSSSNTVDSTNTMVDQSLSSNSDEKASVTSVKGKTRATKGTATDPSLYA 182

128 RRRERISEKIRILKRIVPGAKMDTASMLDEAIRYTKFLKROVRLQ----- 175

183 RKRREKINERLKTQNLVPGTKVDISTMLEAVHYKFLQIQIKVRLKGFVFVSQNL 242

QY 176

176--PHSQI-----GAPMA-NDSYLCYHN 195

243 IITHVQLSSDDLMMYAPFLAYNGIDMGFHHN 273

RESULT 7

T05216

hypothetical protein F1715.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999

C:Accession: T05216

R:Bevan, M.; Vitale, D.; Liguori, R.; Argitton, A.; De Simone, V.; Hohnsbeil, J.; Mewes,

submitted to the Protein Sequence Database, July 1998

A:Accession: T05216

A:Molecule type: DNA

A:Residues: 1-349 <REV>

A:Cross-references: EMBL:AL031032

A:Experimental source: cultivar Columbia; BAC clone F1715

A:Map position: 4

A:Insertions: 177/3; 279/3; 301/3; 323/3

A:Note: F1715.70

Query Match 16.28; Score 170.5; DB 2; Length 349;

Best Local Similarity 44.68; Pred. No. 3.5e-07;

Matches 33; Conservative 20; Mismatches 18; Indels 3; Gaps 1;

QY 104

104 DPATVPKPN---RRNVRISDDPOTVARRRERISERIKIRILKRIVPGAKMDTASMLDEA 160

222 DPSKALNNGTRASRGAAATPQSLYAKRRERINERLRIQNLVPGTKVDISTMLEEA 311

QY 161

161 IRYTFELKROVRL 174

312 VHYVKEFLQIQIKL 325

RESULT 8

G85362

hypothetical protein AT4g30980 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: G85362

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Coli-

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: G85362

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-367 <STO>

A:Cross-references: GB:NC_001268; NID:97270000; PIDN:CAH79816.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g30980

A:Map position: 4

Query Match

Best Local Similarity 16.18; Score 169.5; DB 2; Length 367;

Matches 61; Conservative 24; Mismatches 68; Indels 75; Gaps 9;

QY 22

22 SNHSPKRSME-----PQPHLLMDNKKANDLTOEHAFLNDPNNL----- 63

46 SHSPKRRKTEAMNSSLTTPSSSPH-----LQSPATFDHDDL--HHIFSSRP 94

QY 64

64---MID--PPET-----LIHLDEDE-----EYEDMDAM----- 88

95 WPSVLDTPPTSCAPYTGPHHDASRQITMIPLSHNPNDALFNGFTGSLPFL 154

89 -----KMQQWIAWMPVDIPATVPKPNRRNVRISDDPOTVARRRERISERIK 138

155 POGSGGQTOGQATASATGATGATQOPTRKPVRRARQO-ARDPHSIARLRERIAERH 213

QY 139

139 RILKRIVPGAKMDTASMLDEAIRYTKFLKROVRLQPHSQIAPMAN 186

214 KSLQELVPNGNKTDKASMLDEIIVYKFLQIQVLSMSRGLGAASAS 261

RESULT 9

E84634

hypothetical protein At2g4260 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84634

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tal-

euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ve-

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: E84634

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-350 <STO>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2002, 10:27:54 ; Search time 5365.19 Seconds

(without alignments)
11856.640 Million cell updates/sec

Title: us-09-548-971b-1
Perfect score: 3856
Sequence: 1 cctcagaccatctatctatcc.....caagaggggtccctcctagag 3856

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_tun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htgo_iny: *
32: em_htgo_rtd: *
33: em_htg_hum: *
34: em_htg_iny: *
35: em_htg_rtd: *
36: em_htg_other: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3842	99.6	95643	8 F6N15	AF069299 Arabidop
2	3842	99.6	197119	8 ATCHRY1	AL161471 Arabidop.
3	3114	80.8	167234	8 AC004708	AC004708 C15B11.
4	324.6	8.4	367	8 AY027278	AY027278 Arabidops
5	172	4.5	68098	8 AB020752	AB020752 Arabidops
6	172	4.5	92491	8 ATE17114	AL135394 Arabidops
7	133.2	3.5	90448	8 AC079733	AC079733 Arabidops
8	133.2	3.5	90448	24 AC079730	AC079730 Arabidop
9	129.8	3.4	187576	2 AL139811	AL139811 Homo sapi
10	128.4	3.3	14867	2 AE001398	AE001398 Plasmodu
11	128.2	3.3	987734	2 PFMAL1P2	AL031745 Plasmodu
12	125.2	3.2	4601	3 DMU11584	U11584 Drosophila
13	125.2	3.2	19517	3 DMU37541	U37541 Drosophila
14	125.2	3.2	74342	8 AB020742	AB020742 Arabidops
15	124.8	3.2	36977	2 AC092304	AC092304 Homo sapi
16	124.2	3.2	176174	9 AC007483	AC007483 Homo sapi
17	122	3.2	175223	3 AC004617	AF205580 Plasmodu
18	121.8	3.2	6598	3 AF205580	AF205580 Plasmodu
19	121.6	3.2	14601	3 DMU11584	U11584 Drosophila
20	121.6	3.2	19517	3 DMU37541	U37541 Drosophila
21	120.6	3.1	126999	9 AL513328	AL513328 Human DNA
22	120.6	3.1	234112	3 PFMAL1P2	AL031745 Plasmodu
23	120.4	3.1	67970	3 PFMAL1P3	AL031746 Plasmodu
24	120.2	3.1	104992	2 AC005504	AC005504 Plasmodu
25	120.2	3.1	111122	8 ATE11C1	AL132976 Arabidops
26	120.2	3.1	169546	2 AC004157	AC004157 Plasmodu
27	120	3.1	160416	2 AL590287	AL590287 Homo sapi
28	117.8	3.1	86827	3 PFMAL1P5	AL034556 Plasmodu
29	117.8	3.1	326924	2 AC093082	AC093082 Homo sapi
30	117.8	3.1	326924	2 AE001398	AE001398 Plasmodu
31	117.6	3.0	14867	3 AE001398	AE001398 Plasmodu
32	117.6	3.0	86827	3 PFMAL1P5	AL034556 Plasmodu
33	117.2	3.0	175223	3 AC004617	AC004617 Homo sapi
34	116.8	3.0	137889	9 AC073269	AC073269 Homo sapi
35	116.6	3.0	126999	9 AL513328	AL513328 Human DNA
36	116	3.0	143331	9 AC091214	AC091214 Homo sapi
37	115.6	3.0	57970	3 PFMAL1P3	AL031746 Plasmodu
38	115	3.0	183584	9 AC012492	AC012492 Homo sapi
39	114.8	3.0	136098	9 AC006970	AC006970 Homo sapi
40	113.8	3.0	163443	2 AC006280	AC006280 Plasmodu
41	113.8	3.0	318221	2 PFMAL1P3	AL049184 Plasmodu
42	113.4	2.9	15421	3 PFCOMP1RA	X95275 P. falciparu
43	112.8	2.9	213530	2 AC090493	AC090493 Mus muscu
44	111.6	2.9	149733	3 AC087568	AC087568 Pan trogl
45	111.2	2.9	6598	3 AF205580	AF205580 Plasmodu

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	REFERENCE	AUTHORS	JOURNAL	TITLE
1	F6N15/c	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
2	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
3	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
4	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
5	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
6	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
7	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
8	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
9	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
10	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
11	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
12	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
13	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
14	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
15	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
16	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
17	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
18	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
19	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
20	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
21	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
22	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
23	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
24	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
25	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
26	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
27	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
28	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
29	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
30	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
31	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
32	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
33	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
34	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
35	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene
36	F6N15	Arabidopsis thaliana BAC F6N15.	AF069299	1	GT:3193311	Arabidopsis thaliana	Stratagene	Stratagene	Stratagene	Stratagene	Stratagene

REFERENCE 3 (bases 1 to 95643)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAY-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63108, USA
 e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, Mashu, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 3' cosmid is F5110, 1300 bp overlap. Actual start of this cosmid is at base position 1 of F6N15, actual end is at 1950 F5110.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneIndex (P. Green and L. Hillier, ms in preparation).

FEATURES

Location/Qualifiers

1..95643

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/clone="F6N15"

/chromosome="IV"

/map="unknown"

/complement(829..4226)

/gene="F6N15.14"

/complement(join(829..942,1006..1086,1156..1237,1315..1500,1574..1675,1776..1809,1863..2001,2106..2230,2325..3299,3548..3709,3844..3987,4073..4226))

/gene="F6N15.14"

/note="contains similarity to breast cancer susceptibility (Brca2)"

/codon_start=1

/evidence-not_experimental

/protein_id="AAC19315.1"

/db_xref="GI:3193333"

/translation="MSTQQLPDDSGDGFRRFWARIIQSYSDSTPTKALESTAPLS

MDLLIQCCKSLAREAMPGETPERRTGLGKSVYLKSAKASIIAEVITSDR

NTMCSIPOHROVDATLTLPFRPTASGKSVPLKESITAKMSIISDKITIDSNVLPPE

SGFVNSLFOIASNKNVNSAGLARAKALGLEEDLNGFNHYNSSSSSQOHGS

GLKTHEPDATVVKHSGTGPQYEDYVSGKSEVNPGLKVPPTFOAGKSGSVA

EALKRANRLGDELGSFFDVGDDOFTPEKDRISDAIINNGSARGYIAEERT

SNKRTNSFVSLMSSSKOFSSVNLNLASGCLIKKPDVAVDEPDCLNRTGLSNK

RLSASPMAYNSKNGFIPIRGROGRPADQPIYDITNRDRIAYANKODSTQKRLK

TVSVPFKPRISSEKTPSKHALDASSVSCDITLSKRVASTRIPESPVYIK

DFGGMPTATITRPMSESNQVQIMVYMSVLSLQTRLEKLFKCMISLTPYVMBE

KKVCDSFEACMIYMKLACDIYIPACRGNFLLITVLELREYREYREYVMBE

IKKILSGDAPASPMVLCISAINPKTIDNSOFAHCSDCSNVYELTGTGPTSMNAAD

VVLTQUNAGKLFVQKLRHAYVILSYLKIIGAGLSGMAVTPSPLESTIQNPQAVIS

STICLLNNGTIRAHMDRLGFC"

gene 7427..8228
 CDS /gene="F6N15.13"
 join(7427..7890,7976..8041,8120..8228)
 /gene="F6N15.13"

gene 8515..9759
 CDS /gene="F6N15.12"
 join(8515..8695,8783..9759)
 /gene="F6N15.12"

gene 11756..13710
 CDS /gene="F6N15.11"
 join(11756..11816,11897..12294,12667..12732,13158..13624,13710)
 /gene="F6N15.11"
 /note="contains similarity to transcriptional activators such as Ra-1-like and myc-like regulatory R proteins"

gene 15299..20892
 CDS /gene="F6N15.10"
 join(15299..15951,16806..17061,17136..17246,17320..19156,19234,19332..19628,19799..20001,20363..20614,20746,20892)
 /gene="F6N15.10"

gene 20746..20892
 CDS /gene="F6N15.10"
 join(20746..20892)
 /gene="F6N15.10"

/gene="AT4g00020"
/number=4
intron complement(6639, .6711)
/number=4
exon complement(6712, .6813)
/gene="AT4g00020"
/number=5
intron complement(6814, .6913)
/number=5
exon complement(6914, .6947)
/gene="AT4g00020"
/number=6
intron complement(6948, .7000)
/number=6
exon complement(7001, .7139)
/gene="AT4g00020"
/number=7
intron complement(7140, .7243)
/number=7
exon complement(7244, .7368)
/gene="AT4g00020"
/number=8
intron complement(7369, .7462)
/number=8
exon complement(7463, .8437)
/gene="AT4g00020"
/number=9
intron complement(8438, .8685)
/number=9
exon complement(8686, .8847)
/gene="AT4g00020"
/number=10
intron complement(8848, .8981)
/number=10
exon complement(8982, .9125)
/gene="AT4g00020"
/number=11
intron complement(9126, .9210)
/number=11
exon complement(9211, .9364)
/gene="AT4g00020"
/number=12
intron complement(9365, .13366)
/gene="AT4g00030"
12565, .13028
exon /gene="AT4g00030"
/number=1
CDS join(12565, .13028,13114, .13179,13258, .13366)
/gene="AT4g00030"
/note="Coded for by A. thaliana CDNA W43412
contains EST gb:W43412"
/codon_start=1
/product="predicted protein of unknown function"
/protein_id="CA80761.1"
/db_xref="GI:7267090"
/translation="MALALSLACSPPLRTRRAGRTSCSIFANPAORAKRLLELI
SEEDRGRTOKDPKRDIVNAIESMTVIGRSSITDDSLSAIWLMLTTEKQLFI
EKAGLFGTADVDYLDVINKRLINNVITFPPDGVFVSDIDIASPQVNRFSASV
LRGNMELPLPPFGKMFENVMDGEIRAKDIRGDIYLIVDRAPYMTSEFV"
13029, .13113
intron /gene="AT4g00030"
/number=1
exon 13114, .13179
/gene="AT4g00030"
/number=2
intron 13180, .13257

Query Match 99.68; Score 3842; DB 8; Length 197119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ccactactatccggtgtgaccttaagactttgaagactactagaataatgcaaat 67
|||||

Db 44954 CCATCTACTATCCGTTGTGACCCCTTAAACCTTTTGACACTACTAGAAATATGCCAAT 44895
QY 68 aacatgtccatcatccatcccttcttctgtttgaagactgaattcttgaaga 127
|||||
Db 44894 AOCATATGTCATATCCATCCCTTTCTTTGTTGAACTGAACATCTTAATTTGTA 44835
QY 128 gaaaaaaccttatgttaatalcacccgtagaagaaaaaatalatccatcatlaattc 187
|||||
Db 44834 GAAAAAACCTTATGTATATACACCGTAGCCAAAAAATATCTCATATTAATTTT 44775
QY 168 tatataagattatcacctccctgtgtgaagacttaacatccatccatgtgtgatta 247
|||||
Db 44774 TATATAGATTAATACATTCCTCTGTTGTAGACTTACCTCAATTCGACAGTGTGATTA 44715
QY 248 actataaagaagcgaagaagtaggaagcttaataattgatgttgcagtagactgta 307
|||||
Db 44714 ACTAATMAAAGGACGAAGAGTCGACCTTATATATATGATGTGCAATGACTGTA 44655
QY 308 tatgttgatgaatataaagaatgtagaatgaatgaacggggtatttgccttg 367
|||||
Db 44654 TATGTTGATGAATATTAACAGATGACATTAATGATGATAAACGGGATTTTGTCTTG 44595
QY 368 aactccattaaagcgaatgtgaagaagatgtgaggtctaccttgaagaacttaccct 427
|||||
Db 44594 AACTCATTAAGGCAATGTGAAAAAGAGATGTGAGGTCCATTTTGAATAATTATCTCT 44535
QY 428 agcttgcgactttaaactctatgaagaatgaacgacacatataagaattcatgtgagaa 487
|||||
Db 44534 AGCTTTGTCGATTTTAAATCTATGAATGAACCAACATATTAATAATTTGATGTGACAA 44475
QY 488 cgacatttaagcggatcccttaattgaacggaattatgaatatactatataataat 547
|||||
Db 44474 CCACATTTAGACGGTATCTTAATATAGCCGATTAATTAATTAATTAATTAATTAAT 44415
QY 548 taggtgcgaattatgaattctatccactgaagaattgaagaagggcaacttga 607
|||||
Db 44414 TAGTGCTATATTAAGTTTACTTACCATTTGACAAATTAACAAATGSGAATGCTTAA 44355
QY 608 tgcgaaagaagccgtcccaacttcgtgtaatagttatgqgqaaagaaactcgttaaa 667
|||||
Db 44354 TGTCMAAAGAGCCGTCGCCACTCGTGAATGATTAATGSGGAGAGATCCGTAAAT 44295
QY 668 cgtcaataaagaacttaagaactgaagaattgaacccaataatcctaagaagaagcttg 727
|||||
Db 44294 CGTCAATTAACACACTTAAGACCTGAAGAAATGACCAAAATTCATAAGAAAGAGTG 44235
QY 728 aagaagtcattatcgtatccagctcatatctcagctgaagaatcaatcaagccgttg 787
|||||
Db 44234 AAGAAGTCATTTATCGTATCCAGCTCATATTTCCTAGCTAAGATCAATCAAGCCGTG 44175
QY 788 aagggccttgaagaagaatgtcgaagaagccgtgaggtttgaagaagaagaagaatag 847
|||||
Db 44174 AAGGCGCTTGTAACAAATATGCGAAGAACCGGGGTTTGAAGAAAGCAAGAAATAG 44115
QY 848 aagaacaatgactgttaattgacctatcttggtatagaagtgtcacaagaagaagaag 907
|||||
Db 44114 AAGAAACATGATGTTAAATGCTATTTTGTATAGAGAGTTGTCAAAAAGAGAGAGAG 44055
QY 908 aagaagaataggtcaaaaatagagcactaaagaatgagaaatgtgttgtaactat 967
|||||
Db 44054 AAGAGAAATTTAGGTCAAAATTAAGACGACTAAAAATGCGACATGTGTTGAAATACAT 43995
QY 968 tacaagaagcgaattatgcttccctatgagcaatgatatccaaacaaagtgcgaagctct 1027
|||||
Db 43994 TACAAGAGCGACTATGCTTCCTTATGAGCAATATCCAAACAAAGGCAAGCGTCTCT 43935
QY 1028 ttltgcccctaatltgttaagaagctctcctcctcctgcgcctccttagaagaacccctaga 1087
|||||
Db 43934 TTTTGGCCCTTAATTTGTAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 43875
QY 1088 ttatccctgttctgtgactgtccttcttgaagtaacatgatttgcacacacactt 1147
|||||
Db 43874 TTTAATCCCTTGTCTGTGACTCTCTTTTGAAGTAAACCATGATTTTGACCAACACTATT 43815

QY 1148 tctctatccttctgtgtcctatagagatttgccttataatgttcttctgtatgtcctgt 1207
 Db 43814 tctctatccttctgtgtcctatagagatttgccttataatgttcttctgtatgtcctgt 43755
 QY 1208 agctgcgtatcagaaatttaattgttataaagaatttataaactagacaaatgagt 1267
 Db 43754 acgtacgtatcagaaattttaaattgttataaagaatttataaactagacaaatgagt 43695
 QY 1268 ccataaattctgtatagcgaanaagtagaataatatactgaagctttaaagccataata 1327
 Db 43694 ccataaatttctgtatagcgaanaagtagaataatatactgaagctttaaagccataata 43635
 QY 1328 taatttgcgtcgaanaacttagctttaaataatgataatgataatcttcttaactcct 1387
 Db 43634 taatttgcgtcgaanaacttagctttaaataatgataatgataatcttcttaactcct 43575
 QY 1388 tgcgaataatcacatgcaggaatttctgtcgaagtagtgcgcgaattcttcacattcac 1447
 Db 43574 tgcgaataattcacatgcaggaatttctgtcgaagtagtgcgcgaattcttcacattcac 43515
 QY 1448 acacacaaacttaacttagaactccatagaaatctttaaatagacaacttctcgaanaa 1507
 Db 43514 acacacaaacttaacttagaactccatagaaatctttaaatagacaacttctcgaanaa 43455
 QY 1508 aaatttagtatagaanaaacttccagaattaaacatgagctataataattaaataaat 1567
 Db 43454 aaatttagtatagaanaaacttccagaattaaacatgagctataataattaaataaat 43395
 QY 1568 gaagtaataatgataatgataatgataataatcgcgaatgcggtgaanaaataacataatcgt 1627
 Db 43394 gaagtaataatgataatgataatgataataatcgcgaatgcggtgaanaaataacataatcgt 43335
 QY 1628 attaaattgtcatgcagcccatgctcatatgataatgcgaacttcttaattcttaacaaat 1687
 Db 43334 attaaattgtcatgcagcccatgctcatatgataatgcgaacttcttaattcttaacaaat 43275
 QY 1688 aactaataataatcgaanaaactgttctgcgaattaaatcttcttctgtgtcgaact 1747
 Db 43274 aactaataataatcgaanaaactgttctgcgaattaaatcttcttctgtgtcgaact 43215
 QY 1748 tctcaacacttaacagatttgaatgtctctctcttataattcttcttcttcttcttctt 1807
 Db 43214 tctcaacacttaacagatttgaatgtctctctcttataattcttcttcttcttcttctt 43155
 QY 1808 cctacgtcttcttattatgaatcttaagaacgaatgcgtcgaacttctctctctctct 1867
 Db 43154 cctacgtcttcttattatgaatctttaaagacgaatgcgtcgaacttctctctctctct 43095
 QY 1868 tcttcttaagaatttcttgaactcatctctcccaattgaagcagctataatcataatg 1927
 Db 43094 tcttcttaagaatttcttgaactcatctctcccaattgaagcagctataatcataatg 43035
 QY 1928 ttactcttataataatgtcgaattgaataaanaaataaataactgcgtgcgtatctta 1987
 Db 43034 ttactcttataataatgtcgaattgaataaanaaataaataactgcgtgcgtatctta 42975
 QY 1988 ctaatgtctgttgaanaaactcgaatataatgaatcgaatcgaatccttcttataatgataat 2047
 Db 42974 ctaatgtctgttgaanaaactcgaatataatgaatcgaatcgaatccttcttataatgataat 42915
 QY 2048 gtgaggttaaatcttaattcttatacaataaagaacgtgaatgtatgtataltgag 2107
 Db 42914 gtgaggttaaatcttaattcttatacaataaagaacgtgaatgtatgtataltgag 42855
 QY 2108 gaatcagatataatgctgtgtaactatacttgcgaataactggaagatttaactattt 2167
 Db 42854 gaatcagatataatgctgtgtaactatacttgcgaataactggaagatttaactattt 42795
 QY 2168 gcaaaataataatctagctccggaanaaatacagaacgagcagcagcagcagcagcagcag 2227
 Db 42794 gcaaaataataatctagctccggaanaaatacagaacgagcagcagcagcagcagcagcag 42735

QY 2228 gttgaaatcttgggaanaatttgtatatacaataacatataataaataatgattctgtgtgag 2287
 Db 42734 gttgaaatcttgggaanaatttgtatatacaataacatataataaataatgattctgtgtgag 42675
 QY 2288 ttcaaaagccttctcaaaatctcaattctttaaactcgaagacattttacatagaanaata 2347
 Db 42674 ttcaaaagccttctcaaaatctcaattctttaaactcgaagacattttacatagaanaata 42615
 QY 2348 agttgaatcataaanaataatgatttatttcttgaaggtcttcttcttcttcttcttcttctt 2407
 Db 42614 agttgaatcataaanaataatgatttatttcttgaaggtcttcttcttcttcttcttcttctt 42555
 QY 2408 taataagaagttgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2467
 Db 42554 taataagaagttgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 42495
 QY 2468 ttcatnaaacatataatlaagagatacaatlaatacaanaataatcttaataacttaaa 2527
 Db 42494 ttcatnaaacatataatlaagagatacaatlaatacaanaataatcttaataacttaaa 42435
 QY 2528 ttacataataaatttcaacaacatctuaacttaattgaacagaanaagtaaaatgataat 2587
 Db 42434 ttacataataaatttcaacaacatctuaacttaattgaacagaanaagtaaaatgataat 42375
 QY 2588 atcatataaattgaacacataatctcatalaanaaanaaataatgaagcagatgaagtaataata 2647
 Db 42374 atcatataaattgaacacataatctcatalaanaaanaaataatgaagcagatgaagtaataata 42315
 QY 2648 gagacatgcatgtgaagcattcgcgttataatcgcgaatgcgaagataatcagtaataatc 2707
 Db 42314 gagacatgcatgtgaagcattcgcgttataatcgcgaatgcgaagataatcagtaataatc 42255
 QY 2708 atatgtgataatcttctggaanaaagaatatacatatgaagaataaagaagaatgaaatg 2767
 Db 42254 atatgtgataatcttctggaanaaagaatatacatatgaagaataaagaagaatgaaatg 42195
 QY 2768 gaaatcgtatcgtataaanaagaagaagtgatgcgaacttctgtctcgttcccaagaacga 2827
 Db 42194 gaaatcgtatcgtataaanaagaagaagtgatgcgaacttctgtctcgttcccaagaacga 42135
 QY 2828 tccaacacagcccccaagaagaatgagatgcgaactcgaactcgaactcgaactcgaact 2887
 Db 42134 tccaacacagcccccaagaagaatgagatgcgaactcgaactcgaactcgaactcgaact 42075
 QY 2888 tggacaagaacttaatgattcttctcaacaagaacaggaacttctcgaatgactccac 2947
 Db 42074 tggacaagaacttaatgattcttctcaacaagaacaggaacttctcgaatgactccac 42015
 QY 2948 catctcagatgataatcgaactcccggaacctaattcaacttggagcgaagaacgaagatc 3007
 Db 42014 catctcagatgataatcgaactcccggaacctaattcaacttggagcgaagaacgaagatc 41955
 QY 3008 gatgaagaacatgataatgataatgataatgataatgataatgataatgataatgataatg 3067
 Db 41954 gatgaagaacatgataatgataatgataatgataatgataatgataatgataatgataatg 41895
 QY 3068 gacatcgaactcgtcgaacaggtccatgaacgaacgcgcgtgaatgaatgaatgaatgaatg 3127
 Db 41894 gacatcgaactcgtcgaacaggtccatgaacgaacgcgcgtgaatgaatgaatgaatgaatg 41835
 QY 3128 cctcagaacgtgtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 3187
 Db 41834 cctcagaacgtgtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 41775
 QY 3188 aagatgacgt 3247
 Db 41774 aagatgacgt 41715
 QY 3248 cgttaacacaaatccttgaacaggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 3307
 Db 41714 cgttaacacaaatccttgaacaggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 41655
 QY 3308 gctcctatgcttaacccctcttaacttcttattatcaacaaactcccaacccctgaatgaact 3367

COMMENT

(E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935,
 Fax: 81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=MT16
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
<http://combio.ornl.gov/Grail-1.3/>),
 GENSCAN (Chris Burge, MIT, <http://CCR-091.mit.edu/GENSCAN.html>),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and
 SplicePredictor (Volker Brendel, Stanford University,
<http://gremlini1.zool.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-se/>).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MT19.

FEATURES

Source

CDS

CDS

CDS

Location/Qualifiers
 1..68098
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MT16"
 /clone_1lb="Mitsui P1"
 join(<1..157,630..706,796..868,974..1330)
 /note="gene_id:MT16.1
 pir1|T05407
 strong similarity to unknown protein"
 /codon_start=2
 /evidence=not_experimental
 /protein_id="BAB17027.1"
 /db_xref="GI:11022654"
 /translation="IPPLAVFNRHAYGRSLAENSHLQSMATALFRSIRRDVYS
 APLSVKSLAGNAQPSMGSSVIGQVASFSPGSKPVVNDLGTGTNNALREBE
 KSKSTEAMVGAOLTRSPFALDVGTSRFTSTIGDJKTQDEPKIKSPFPLPHLSV
 YPOKNSMLSTFNRISGYLTGYTFAGLILYLMKMGICLTPSPYOVLYHTCOQLYVI
 TSTPALAIYHRIKSTHSLTH"
 2019..3572
 /note="gene_id:MT16.2
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09516.1"
 /db_xref="GI:9758989"
 /translation="MADONGKNPFRNDGTYTLTATNNLKLIVANYNHQEGESGNTN
 IASNNHPPPLTAFAKSLITANKNOEPESGNTTINRKNHPNPPLTAFNNLNL
 YANYNHQEGESGNTNINNNNDNPPLTAFNNLIVANYNHQEGESSNNYNPLNLY
 NVGHISAASFNAFPPTPSLTPQPDSSRSYSGKPPPLSFVSPGVADKRMNLSGL
 DMKTAORTFEROKYLDLDYPTAERSHSLFGSALTTTKRFLFATNNGSGAL
 RLFRSPSLDLFLCAVDNFPFLMSDKYGRGLIPAIRAVDKTKESLYKLTLYET
 LHLARLEGCLALNNVLOEIGIYRDLFECAVANNADMLSPDYGTHVGNLILNP
 VATTAIARLRCGSPFLAMEROGSYVYKCKKSPAPADQVLEERGNAKKMYMTDK
 PGNFYQSLNLRMKEMKPLRPREVKLRHFGKMEIGKORNTLRYIODEIVGMINQ
 LPDKSGYVA"
 complement(4620..6215)
 /note="gb|AA014519.1
 gene_id:MT16.3
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09517.1"
 /db_xref="GI:9758990"
 /translation="MDKFSYNSYSDASSPRSRDVEFNPSPWEDQOQNFVKILMC
 SYGKGIORPHNDQTLVYNDCTKINSVDRGIRPALSKSAVCSGGDCGEISFKYQ
 LPEGDLDALISYINDEDELMHHEHYDRLLRSTKPAKMLFLPPSSISGSGFSEST
 KSDPTLNDIPRSEKSVATAPNNADFLFGSKVAPIDPSPKYKQPPPPVYLED
 PGMFYDQMLQREHGVNPAEIORQIQEOMIQTINDQOQMLHOMQLHQOQOQDEA:HQ

CDS

CDS

CDS

tRNA

CDS

NQLHQOEVIHONOLLQOEAIHONOMLQOQOQOQEAAMYRRKTEDEAGRFFPYTQ
 PVTNQPVPVGMQNTNNSVJOCNIYTTISQNLPEQOQOQOQYMPAPSOAPGT
 SMRPTVOGNOGYRPSPPVORLHHPDAVMEQONPGVNVQPPPTSGGQVNTSV
 VMTSGVPMLOEYFSQMKPVYHTYTAGDEGMWQPPAPQOQOQOQOQMPVS
 DLRTGPRGKAVYNNAPQVSDSV"
 complement(6879..8039)
 /note="contains similarity to unknown protein
 gb|AA015577.1
 gene_id:MT16.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09518.1"
 /db_xref="GI:9758991"
 /translation="MDVTGTVTRDAPFDRYSKKQKLYHSVTDVLCVCGTODI
 IOLNDQGVPESEVTELRKRLDALLPITOLQSKNETKWSLSKLKLLVSYEP
 LACFVSQINLVNKKIILHHCYREGDVFQDQVKEAGEFEIEVLSQFLEFHQI
 LKLNIEFPAMWIIANGKSKSSKLEFLSLKQCDILREKSDALEVARTH
 YLHKEIKOKLITGLMIGNEKSPYAEIYSQCMKVKTEKLEMEYHHLDDPIN
 KVALSAGEYSPLIKLVHMLAIKQEQMOKLIPCELTGNEYKFEHSAFVCPVS
 SEENPQMQLPCGVNISKQSMKLSKNCARTKCYCTARETASACRQLYE"
 101419236..9366,9468..9613,9767..9843,9936..10058,
 10157..10259,10444..10571,10659..10745,10831..10929
 11073..11152,11214..11327,11415..11531,11615..11660
 11749..11829)
 /note="gene_id:MT16.5"
 /codon_start=1
 /evidence=not_experimental
 /product="serine carboxypeptidase"
 /protein_id="BAB09519.1"
 /db_xref="GI:9758992"
 /translation="MRNLSEIVLFLTLFPIHHLVDASLVLKLPFEGPFLFEL
 YVSGESGOVELFYVFKSERENDEPMLMLTGCGSCSSICGLLEPFLFKGI
 NGYAPPELTSFMTKANYLILEAPAGSGVYAKTRAPFESSPTKQMOIDQFII
 FVKRPFTSNPFYGGSGYSCKYVRANQVISTVTSYILIDQYVYGNVYTKNIT
 IRVFAHGMGLISDLEFSELRSCGGFHVDPDSNRCNSRQADYTHQSELYSEH
 LKRCQVYVLADPTNITRDRRTYRFLSAFMANDESNRRLGVKVPYQKNNKCN
 NIPYTFEFNAVVPYVNNLSKGRSLYSQSDHDSMSPFSQTMKIALYVSDOK
 MMSSNOVAGYTRANKTFEATIKGGGHTAEVTPYQCSLMFRMWIDCEPL"
 complement(12104..12286,12381..12479,12565..1274
 12836..12936,13097..13257,13714..13980))
 /note="emb|CA89165.1
 gene_id:MT16.6"
 /codon_start=1
 /evidence=not_experimental
 /product="inorganic pyrophosphatase-like protein"
 /protein_id="BAB09520.1"
 /db_xref="GI:9758993"
 /translation="MAATRYLTATAVTOTSCFLAKQAFILPARKSGGFGGLCF
 RALYKSRPESGSAIYNPOVKYQEPKASLDYRPFLDSSKKYKSPEDIRTL
 GVNFPIYEIRKESAKKEVATIDEDFPPIKDKKGLRYTPYNNNYGLLQYWE
 SHASEVOCFPGDNDPVDAVEIGETQKIDDTLTKIPPLAALAMIDGELDKATVAI
 DDPAKHLVNOVENYKHPVTLTAIRMPFVDYKIPPOKPAKRFGLDCKPANKDYAL
 IOETNESMAVLKYSVAGLSLY"
 14190..14266
 /note="gene_id:MT16.7"
 /evidence=not_experimental
 /product="tRNA-Val(AC)"
 complement(14508..14580,14673..14767,14841..1493
 15029..15148,15236..15325,15411..15725,15814..16132,
 16352..16414)
 /note="emb|CA10321.1
 gene_id:MT16.8"
 /codon_start=1
 /evidence=not_experimental
 /product="mitochondrion NAD-dependent malate dehydrogenase"
 /protein_id="BAB09521.1"
 /db_xref="GI:9758994"
 /translation="MEFRGDANQIRAIRISNHLTPQEMANKSVYGRNCRKAGNCP
 VALIAGAGIGOSISLIMKNNPLVSLHLVDVYVAGVYADVSHMTGAVVGGFGL
 QLEALIMQDLYIPAIIPKPMPTDRDLKINAGIVKTLCEGVAKCAINAVNLIS
 PVNSTVLAIEVKKATIPKPKLITVTTLDAVRAATPAEVLGIDPREVDVYVQ
 AGVITLLELSQVPPSPFTQETELTNRLQNGTIVAEKKGAGSATISKAYAAK
 ADACLRGLRQDANVCEFSVASQVTELAIPATVAVKLRGKGAEEVYOLGFLNEYERIC


```

exon          8551..8780
              /gene="F17114_20"
              /number=3
              /gene="F17114_20"
              /number=3
              /gene="F17114_20"
              /number=4
              /gene="F17114_20"
              /number=4
              /gene="F17114_20"
              /number=5
              /gene="F17114_20"
              /number=5
              /number=6
              /gene="F17114_30"
              /note="similarity to predicted proteins, Arabidopsis
              thaliana"
              /codon_start=1
              /product="putative protein"
              /protein_id="CAB89352.1"
              /db_xref="GI:7671411"
              /translation="MASNPDLTMCCKEERKOEFFKRLKSDMSSEDTRAIIPYCVINF
              SNEISLGMKPEFVONGSMWKYISKNPFFYEMKSGMEKFEVINALGDHETITPHKG
              AMSGFWKLEFNKDGKEMQPPQSRASEASSKQEDQKREPELVSSDSRSRPTNA
              AATNAGSGYKRRKLNGKKAETOTYKRTENTQNSKRTERYVSKERYIAGPSSVAG
              RFIISKSTIKSLAIPKPFNGNMPKRTIRVIHPHPDGTKVAVVFAKERQDIFSQGM
              KRLCEYEPVEDTCKFTLTPELLVAVSK"
              /number=1
              /gene="F17114_30"
              /number=1
              /gene="F17114_30"
              /number=1
              /gene="F17114_30"
              /number=2
              /number=2
              /gene="F17114_30"
              /number=2
              /number=3
              /gene="F17114_30"
              /number=3
              /number=4
              /gene="F17114_40"
              /number=4
              /gene="F17114_40"
              /number=5
              /gene="F17114_40"
              /number=5
              /number=6
              /gene="F17114_40"
              /number=6
              /number=7
              /gene="F17114_40"
              /number=7
              /number=8
              /gene="F17114_40"
              /number=8
              /number=9
              /gene="F17114_40"
              /number=9
              /number=10
              /gene="F17114_40"
              /number=10
              /number=11
              /gene="F17114_40"
              /number=11
              /number=12
              /gene="F17114_40"
              /number=12
              /number=13
              /gene="F17114_40"
              /number=13
              /number=14
              /gene="F17114_40"
              /number=14
              /number=15
              /gene="F17114_40"
              /number=15
              /number=16
              /gene="F17114_40"
              /number=16
              /number=17
              /gene="F17114_40"
              /number=17
              /number=18
              /gene="F17114_40"
              /number=18
              /number=19
              /gene="F17114_40"
              /number=19
              /number=20
              /gene="F17114_40"
              /number=20
              /number=21
              /gene="F17114_40"
              /number=21
              /number=22
              /gene="F17114_40"
              /number=22
              /number=23
              /gene="F17114_40"
              /number=23
              /number=24
              /gene="F17114_40"
              /number=24
              /number=25
              /gene="F17114_40"
              /number=25
              /number=26
              /gene="F17114_40"
              /number=26
              /number=27
              /gene="F17114_40"
              /number=27
              /number=28
              /gene="F17114_40"
              /number=28
              /number=29
              /gene="F17114_40"
              /number=29
              /number=30
              /gene="F17114_40"
              /number=30
              /number=31
              /gene="F17114_40"
              /number=31
              /number=32
              /gene="F17114_40"
              /number=32
              /number=33
              /gene="F17114_40"
              /number=33
              /number=34
              /gene="F17114_40"
              /number=34
              /number=35
              /gene="F17114_40"
              /number=35
              /number=36
              /gene="F17114_40"
              /number=36
              /number=37
              /gene="F17114_40"
              /number=37
              /number=38
              /gene="F17114_40"
              /number=38
              /number=39
              /gene="F17114_40"
              /number=39
              /number=40
              /gene="F17114_40"
              /number=40
              /number=41
              /gene="F17114_40"
              /number=41
              /number=42
              /gene="F17114_40"
              /number=42
              /number=43
              /gene="F17114_40"
              /number=43
              /number=44
              /gene="F17114_40"
              /number=44
              /number=45
              /gene="F17114_40"
              /number=45
              /number=46
              /gene="F17114_40"
              /number=46
              /number=47
              /gene="F17114_40"
              /number=47
              /number=48
              /gene="F17114_40"
              /number=48
              /number=49
              /gene="F17114_40"
              /number=49
              /number=50
              /gene="F17114_40"
              /number=50
              /number=51
              /gene="F17114_40"
              /number=51
              /number=52
              /gene="F17114_40"
              /number=52
              /number=53
              /gene="F17114_40"
              /number=53
              /number=54
              /gene="F17114_40"
              /number=54
              /number=55
              /gene="F17114_40"
              /number=55
              /number=56
              /gene="F17114_40"
              /number=56
              /number=57
              /gene="F17114_40"
              /number=57
              /number=58
              /gene="F17114_40"
              /number=58
              /number=59
              /gene="F17114_40"
              /number=59
              /number=60
              /gene="F17114_40"
              /number=60
              /number=61
              /gene="F17114_40"
              /number=61
              /number=62
              /gene="F17114_40"
              /number=62
              /number=63
              /gene="F17114_40"
              /number=63
              /number=64
              /gene="F17114_40"
              /number=64
              /number=65
              /gene="F17114_40"
              /number=65
              /number=66
              /gene="F17114_40"
              /number=66
              /number=67
              /gene="F17114_40"
              /number=67
              /number=68
              /gene="F17114_40"
              /number=68
              /number=69
              /gene="F17114_40"
              /number=69
              /number=70
              /gene="F17114_40"
              /number=70
              /number=71
              /gene="F17114_40"
              /number=71
              /number=72
              /gene="F17114_40"
              /number=72
              /number=73
              /gene="F17114_40"
              /number=73
              /number=74
              /gene="F17114_40"
              /number=74
              /number=75
              /gene="F17114_40"
              /number=75
              /number=76
              /gene="F17114_40"
              /number=76
              /number=77
              /gene="F17114_40"
              /number=77
              /number=78
              /gene="F17114_40"
              /number=78
              /number=79
              /gene="F17114_40"
              /number=79
              /number=80
              /gene="F17114_40"
              /number=80
              /number=81
              /gene="F17114_40"
              /number=81
              /number=82
              /gene="F17114_40"
              /number=82
              /number=83
              /gene="F17114_40"
              /number=83
              /number=84
              /gene="F17114_40"
              /number=84
              /number=85
              /gene="F17114_40"
              /number=85
              /number=86
              /gene="F17114_40"
              /number=86
              /number=87
              /gene="F17114_40"
              /number=87
              /number=88
              /gene="F17114_40"
              /number=88
              /number=89
              /gene="F17114_40"
              /number=89
              /number=90
              /gene="F17114_40"
              /number=90
              /number=91
              /gene="F17114_40"
              /number=91
              /number=92
              /gene="F17114_40"
              /number=92
              /number=93
              /gene="F17114_40"
              /number=93
              /number=94
              /gene="F17114_40"
              /number=94
              /number=95
              /gene="F17114_40"
              /number=95
              /number=96
              /gene="F17114_40"
              /number=96
              /number=97
              /gene="F17114_40"
              /number=97
              /number=98
              /gene="F17114_40"
              /number=98
              /number=99
              /gene="F17114_40"
              /number=99
              /number=100
              /gene="F17114_40"
              /number=100
              /number=101
              /gene="F17114_40"
              /number=101
              /number=102
              /gene="F17114_40"
              /number=102
              /number=103
              /gene="F17114_40"
              /number=103
              /number=104
              /gene="F17114_40"
              /number=104
              /number=105
              /gene="F17114_40"
              /number=105
              /number=106
              /gene="F17114_40"
              /number=106
              /number=107
              /gene="F17114_40"
              /number=107
              /number=108
              /gene="F17114_40"
              /number=108
              /number=109
              /gene="F17114_40"
              /number=109
              /number=110
              /gene="F17114_40"
              /number=110
              /number=111
              /gene="F17114_40"
              /number=111
              /number=112
              /gene="F17114_40"
              /number=112
              /number=113
              /gene="F17114_40"
              /number=113
              /number=114
              /gene="F17114_40"
              /number=114
              /number=115
              /gene="F17114_40"
              /number=115
              /number=116
              /gene="F17114_40"
              /number=116
              /number=117
              /gene="F17114_40"
              /number=117
              /number=118
              /gene="F17114_40"
              /number=118
              /number=119
              /gene="F17114_40"
              /number=119
              /number=120
              /gene="F17114_40"
              /number=120
              /number=121
              /gene="F17114_40"
              /number=121
              /number=122
              /gene="F17114_40"
              /number=122
              /number=123
              /gene="F17114_40"
              /number=123
              /number=124
              /gene="F17114_40"
              /number=12
```

Query Match	4.5%	Score 172:	DB 8:	Length 92491:
Best Local Similarity	71.5%	Pred. No. 5.3e-26:		
Matches 226:	Conservative	0:	Mismatches 90:	Indels 0:
Gaps				
Query	2970	ccggaacccttaattccatctggcgcggaagcgcgaagatlaacgatgaagacatgga:gcg#ata	3029	
DB	21872	CCGGGATCAGGAAGACGACGAGGAAGGAAGAACCTCTAGAGGAACCTCGTGTCTATGA	21933	
QY	3030	aggaqatgcacgtacatgcacgcgcgtcatgcagccrctgtagacatcgaccctgcgcagatcc	3085	
DB	21932	AGGAATGATGATGATCAAGATGCGGACCATCATCGTGTGACATGACCCACACCTCTCA	21951	
QY	3090	ctaagccgaaccgcgcatacgaatgaagataaagcgaacatccctcagaacgttgatgctgcgc	3149	
DB	21992	AGAAACCCAAACCCCGTAAGCTGAGGATCTCCGACGACCCCTCAAGATGTGGCGCTAGAC	22055	
QY	3150	ggcgtcggaagaagatcagcgaagaagatccgaatctcaacaaagagatccgtgacctgagt	3209	
DB	22052	ATGCGCGTGAAGAGATCAGTGAAGAGATGAGATCTTCCAGAGACTGCTGCCAAGCTGCA	22111	
QY	3210	cgaagatgcagacagcttccatgcgtcgcgaagacacatacgttacaacaaagtctctgaaac	3269	
DB	22112	CTAAATGATGATGCGCTTCAATCTCGATGAGAGCTAATGCGTATGCTCAAGTTCTTGAAC	22177	
QY	3270	ggcaggtgagatcttct 3285		
DB	22172	GGCACATCGCGCTTACT 22187		


```

CDS
    join(12799..12984,13085..13522,13608..14039,14131..14580,
14676..14927)
    /gene="T8L23.4"
    /codon_start=1
    /product="jasmonate inducible protein, putative"
    /protein_id="AG50741.1"
    /db_xref="GI:12321345"
    /translation="MEVIGSTEGYTFDDSDHDYDKIVGGGROGTHIEFVYKNG
QLESGVHLGVRKGFTEFEINHNNEHESVGGITDYGSGITDGLQETKFRVSELI
GIDEGTFFSLVSGKRIIGFHGYMKERKISLISGYSWIHPKMEKSGKGGQMDG
TNDGVYRKIHVGGVEGIIQIKEDYVRSQKHNSIHGSGSGTQIEFIDJNEH
LVCEGYIDDESGVICALQETNITKSELGKYKKKSLVDKRIYGFHJAKNL
NSLCAYFTVSPKSECGSSKGIYMDQYDFDEIFVYSSNMVNRKIKFHYANAV
VVRQHGNSIVIEEDGEKEPEFLDPNEITSEVCEKMSFSESRISLTPKSKGTS
PTIGASGTRFLASKCAVYGFYGRHDDDLVAIGATFSPLPPPAEKLAQGGGQ
DSMDQVFBGVKRLIVGGKNCVAFKLVVDYDVAIGDHGKNTLFEVKEEYEP
SEYITAVDGCYNKNGTEVEVITMLRIQTKRRTSIPVGCSSNFVLKREGVKIVGFH
GKASMTNIQGVHYVPLE"
    complement(1537..15449)
    /rpt_family="AT rich"
    /rpt_(<15722..16195,16329..>16895)
    /gene="T8L23.5"
    /note="predicted by genemark hmms"
    join(15722..16195,16329..16895)
    /gene="T8L23.5"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AG50744.1"
    /db_xref="GI:12321348"
    /translation="MDTLPMLHLDLIFRIDHSLMIQCTNRSLQSHISKTPNFASE
YFSIRSSLVIATHGCPILCYHVSRSPTVEITQCHILGSGLLLELNKL
CVANPLKRFKFLHNSKRFEPIDFTLRGROOVIGFAVDQRTTSPFKIVINIKGA
VHMLRNDGSTVAFNLKTEKARLIPRPKTRICKITFTVANNLTIISATSEVIYVYA
FHNLSDPKXVLYQINQVILDERKLSWVEYDDKSLVLRBTIIKKDYIKNHLKQV
LHGIDLANKMEVIGSLPEWSPGLDPTQTPSPSSVIGLDDGESEIILACDCKRIPS
VNAIMRLIVGIS"
    complement(18101..18211,18372..18517,18621..18726,
18800..18877,18965..19018,19096..19156,19248..19459,
19540..19613,19700..19829,19912..19994,20071..20269)
    /gene="T8L23.6"
    complement(18101..20269)
    /gene="T8L23.6"
    /note="similar to pectinacetylsterase precursor
GI:1431629 from (Vigna radiata)"
    complement(18101..18211,18372..18517,18621..18726,
18800..18877,18965..19018,19096..19156,19248..19459,
19540..19613,19700..19829,19912..19994,20071..20269)
    /gene="T8L23.6"
    /codon_start=1
    /product="pectinacetylsterase precursor, putative"
    /protein_id="AG50747.1"
    /db_xref="GI:12321351"
    /translation="MKKLMSWILFNIHVGMMMEFDEMEFTVNGTKVQOTQDV
Query Match 3.58: Score 133.2; DB 8: Length 90448;
Best Local Similarity 54.88; Pred. No. 2.2e-10;
Matches 311; Conservative 0; Mismatches 248; Indels 9; Gaps 2;
QY 1407 accgaatttggcaagtatgtgcgaatttcctccatcacacaacctaactaata 1466
DB 55632 AAGAMATATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 65573
QY 1467 gaactctagggaataattttaaataagacaacttatcgaaaaaattagttaga 1556
DB 65572 GCAGTTATGGCCGATCTTAATAGCAAAATTTTGGCTATCCAAATCTTAATA 65513
QY 1527 attccaagaattaaacatga-----gctataataatgaataaataagtaattg 1579
DB 65512 CTAGATGTGATGTCATCAAAATCTTTGGGTCITTTCAAGAAATGATCAAGGCT 65553
QY 1580 aatgataatgaataacatatctgattgcggtaaataaataacatatctgataatgtctc 1639

```

```

DB 65452 CTTGGTGGCAAGTGTCTTCGGAGGTGTCCTCCCTCCCTACACTGGAGATGAAAGTTT 65393
QY 1640 atgcagcccatcactacatgatgacacacgatttattttcacaaataacataata 1699
DB 55392 ATTGGGATTAACAAGAAATATCTCTTAAGAAAGAAATGATCTTAAGCCCTTTGAACAGAT 65333
QY 1700 ttcaaaaaaagatttctcagataaaatttttttgg--tggcawgtcttctcaacct 1757
DB 55332 TCTTGAAGATTAGTCTTGATGATGTCGACATCTTGAAAGAAAGAACCACTTATCAATCT 65273
QY 1758 actaacacagtttgaagatctctctcttatttcttcttctctgattctctatgttt 1817
DB 65272 ACTAGATGATTAAGAAAGCTTTTGCTATATTTCCGCTTTTCGATTTCTTATTTTC 65213
QY 1818 ttattctagaaattttaaagacagatgttttsgtgcgtctctctcttcttcttctctaa 1877
DB 65212 TCGTTATTAATTAATGTAAGCTGAGTGTGTAGTTGCTCTCTGTTTACTGCTTTCGGA 65153
QY 1878 agttactctttgaaactcatctccctcccaattagaacataacataatagttatttta 1937
DB 65152 AATTATATTTTATAGACTATTTCTCTCCAAATTAAGCAATCATGTAAGTTATATTTTA 65093
QY 1938 atatatgtctagttgataaaaaaataga 1965
DB 65092 AATTAATTTCTGCTGATTAATAAAAAAAGA 65065
RESULT 8
AC079730/c standard; DNA; P.N: 90448 BP.
ID AC079730;
AC AC079730;
XX AC079730.1
SV AC079730.1
DT 12-SEP-2000 (Rel. 65, Created)
DT 12-SEP-2000 (Rel. 65, Last updated, Version 1)
DE Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence, complete
DE sequence.
XX HTG.
XX Arabidopsis thaliana (thale cress)
XX Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
XX Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
XX eustosids II; Brassicales; Brassicaceae; Arabidopsis.
XX [1]
XX 1-90448
XX Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
XX Maiti R., Romling C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.,
XX Bowman C.L., White O., Nierman W.C., Fraser C.M.;
XX "Arabidopsis thaliana chromosome 1 BAC T8L23 genomic sequence";
XX unpublished.
XX [2]
XX 1-90448
XX Town C.D., Kaul S.;
XX Submitted (08-SEP-2000) to the EMBL/Genbank/DBJ databases.
XX The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD
XX 20850, USA. cotton@tigr.org
XX
XX SP:REMBL: Q9FD29; Q9FD29.
XX SP:REMBL: Q9FE70; Q9FE70.
XX SP:REMBL: Q9FEV7; Q9FEV7.
XX SP:REMBL: Q9FVS9; Q9FVS9.
XX SP:REMBL: Q9FVS9; Q9FVS9.
XX SP:REMBL: Q9FV70; Q9FV70.
XX SP:REMBL: Q9FV71; Q9FV71.
XX SP:REMBL: Q9FV73; Q9FV73.

```

```
DR SPTREMBL: 09FVT4: 09FVT4.  
DR SPTREMBL: 09FVT5: 09FVT5.  
DR SPTREMBL: 09FVT6: 09FVT6.  
DR SPTREMBL: 09FVT7: 09FVT7.  
DR SPTREMBL: 09FVT8: 09FVT8.  
DR SPTREMBL: 09FVT9: 09FVT9.  
DR SPTREMBL: 09FVU0: 09FVU0.  
DR SPTREMBL: 09FVU1: 09FVU1.  
DR SPTREMBL: 09FVU2: 09FVU2.  
DR SPTREMBL: 09FVU3: 09FVU3.  
DR SPTREMBL: 09FVU4: 09FVU4.  
DR SPTREMBL: 09FVU5: 09FVU5.  
DR SPTREMBL: 09FVU6: 09FVU6.  
DR SPTREMBL: 09FVU7: 09FVU7.  
DR SPTS-PROT: 09FVT2: EPIH_ARATH.  
XX  
CC Address all correspondence to: atc@tigr.org  
CC BAC clone T8L23 is from Arabidopsis thaliana chromosome 1  
CC The orientation of the sequence is from SP6 to T7 end of the BAC  
CC clone.  
CC Genes were identified by a combination of several methods: Gene  
CC prediction programs including Genscan+ (Chris Burge,  
CC http://CCR-081.mt.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,  
CC http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant  
CC of GlimmerM, see Mhaeja Perlea,  
CC http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and  
CC Geneslicer (Mhaeja Perlea and Steven Salzberg, contact  
CC mperlea@tigr.org), searches of the complete sequence against a  
CC peptide database and the plant EST database at TIGR  
CC (http://www.tigr.org/tldb/tgi.shtml). Annotated genes are named to  
CC indicate the level of evidence for their annotation. Genes with  
CC similarity to other proteins are named after the database hits.  
CC Genes without significant peptide similarity but with EST  
CC similarity are named as unknown proteins. Genes without protein  
CC or EST similarity, that are predicted by more than two gene  
CC prediction programs over most of their length are annotated as  
CC hypothetical proteins. Genes encoding tRNAs are predicted by  
CC tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
CC Simple repeats are identified by RepeatMasker (Arian Smil,  
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html).  
XX  
FH Key Location/Qualifiers  
FH  
FH source  
FH 1. 90448  
FH /chromosome="1"  
FH /db_xref="taxon:3702"  
FH /organism="Arabidopsis thaliana"  
FH /cultivar="Columbia"  
FH /map="n1259"  
FH /clone="T8L23"  
FH complement(634..656)  
FH /rpt_family="AT_rich"  
FH join(complement(1711..>1858),complement(<1221..1284))  
FH /gene="T8L23.1"  
FH join(complement(1711..1858),complement(1121..1284))  
FH /codon_start=1  
FH /db_xref="SPTREMBL:09FVU7"  
FH /gene="T8L23.1"  
FH /product="hypothetical protein: 1858-1121"  
FH /protein_id="AAG12106.1"  
FH /translation="MALRKIGWSEGDLMKSDAKPCSRMLKQTAIFTVGALGFVLC  
FH RLHYGPRTVPRSRMAGCGALSVSTALVLFSEPCPNIAAYDNTKTPQAS..P"  
FH join(complement(3378..>3452),complement(<3184..3267))  
FH /gene="T8L23.2"  
FH join(complement(3378..3452),complement(3184..3267))  
FH /codon_start=1  
FH /db_xref="SPTREMBL:09FE70"  
FH /gene="T8L23.2"  
FH /product="hypothetical protein: 3452-3184"  
FH /protein_id="AAG12109.1"  
FH /translation="MGSFLEVLCAIFTPVGVFLRYGLGFVWCILLTLFNFIPGLIY  
FH AIIYLRK"  
FH complement(3452..3549)  
FH repeat_region
```

```
FT  
FT repeat_region  
FT complement(4220..4241)  
FT /rpt_family="AT_rich"  
FT complement(6494..6524)  
FT /rpt_family="AT_rich"  
FT join(<7494..7626,7715..7844,7921..8588,10051..10856)  
FT /gene="T8L23.3"  
FT join(7494..7626,7715..7844,7921..8588,10051..10856)  
FT /codon_start=1  
FT /db_xref="SPTREMBL:09FVU6"  
FT /gene="T8L23.3"  
FT /product="hypothetical protein: 7494-10856"  
FT /protein_id="AAG12112.1"  
FT /translation="MKRNSCTIKQKLRKGLMSEDEKLLNTITKRGHGCWSPV  
FT LERCGKSCRRLWNLNLPDLKRGAFSSSEONLIVELAAVGNMRSQIARLPERTT  
FT KLMNSCIRKKLKKKIDITKPLSEVCKETSRSDNNNSSTFSENDLFEVKK  
FT AEYSAQKESNSVSLRNSLSMTPTGFNIDGVSNAAGDVCVPSIILFPP  
FT SSTVSGODHYNVAEPWMSNSCTSHLNNEMKMSSEYLNES..FSTGVYVKS  
FT NSRIAPNRSQSCDVPKPIQDMARSRFGSGFPPVATTLRRGROVIGFSVQIT  
FT NSRIYIINEVRSNSETTEFEINVGISMKLSFTTLTCTSLNDDMKRFYKKG  
FT LKNDGAIKVAENPETERARLISREPKELOSKTLFTADNNLLISATEEYFVAV  
FT LNDPKWVVKQIRNGVLDKRLYSWPEAYDQCKMLRETLKDKHVKQVHGGDLA  
FT MEVIGSIIPGWTSLDLPFOFTSLSSVIGVDKKEBEIIFADHKIKSSINSLIR  
FT SS"  
FT complement(9315..9339)  
FT /rpt_family="AT_rich"  
FT complement(11306..11348)  
FT /rpt_family="(7)n"  
FT complement(11470..11493)  
FT /rpt_family="(7)n"  
FT join(12799..12984,13085..13522,13608..14039,14131.  
FT 14676..14927)  
FT /gene="T8L23.4"  
FT join(12799..12984,13085..13522,13608..14039,14131.  
FT 14676..14927)  
FT /codon_start=1  
FT /db_xref="SPTREMBL:09FVU5"  
FT /gene="T8L23.4"  
FT /product="hypothetical protein: 12799-14927"  
FT /protein_id="AAG12115.1"  
FT /translation="MEVIGSTEGYTPDGDSDHMDVKKIYVGGRCGIRHYIEFYVW  
FT LESGVHAYRVKGFITFELHNNHLESEGVGYGGLYIOGLQFNTFRYSEL  
FT CEGTKSLSVKGRRIIEFGCYMKERKILSLGQFSMLHPRKMAKSKGQNDQ  
FT DGYTKIHVGCYVHC:OYIKFTVYRKSGOINIGSITGSSGFTQIFETDHLNNEH  
FT EGYVDESQVIOALQFTNKIKSELLGKRGKFSLVDRKRIKVGFGYADKLN  
FT YFTVSPITKFCVGSCKGJWDQGVDFIRIYVSNVNRVIRKFNVRVAVVN  
FT WMSIYDEDEKEFELDPYMLLIISVEGTMSKRSIRLSLFTKSKRTSPTIG  
FT GKRFILASGCAVCGFGRHDDRLVAICAYFSPLEPPPAEKLOAGGNOGSDMD  
FT EGVKRYVGGKNCVAFILKVYVDSNTQVIGEGHKGKTFEVEYVELVPSFYII  
FT CYNKVGTEVEVYITMRITQTKRNTSIPVGCESNSFVLKKEGKIYGFHKKASNM  
FT GAVVPLTE  
FT complement(15377..15449)  
FT /rpt_family="AT_rich"  
FT join(<15722..16195,16329..>16895)  
FT /gene="T8L23.5"  
FT join(15722..16195,16329..16895)  
FT /codon_start=1  
FT /db_xref="SPTREMBL:09FVU4"  
FT /gene="T8L23.5"  
FT /product="hypothetical protein: 15722-16895"  
FT /translation="MDLEPHN:LDLELFRIDHRSIAMTOCTNRSDSHSKSPNFR  
FT FSRIRNSLLIYITHGGLICVHLYGDSRSRPRVEITLTOCHILGYSGLLILKLNK  
FT ANPLTKRFFELNHSRSEKFPPTDFTLRGQOVYIGFVADDTPTBSFKLVNKEG  
FT LKNDSDIVAFNDLKEARLIPRPFKELOSKTLFTADNNLLISATEEYFVAV  
FT LSDPKWVVKQIRNGVLDKRLYSWPEAYDQCKMLRETLKDKHVKQVHGGDLA  
FT LANKWEVIGSIPEWSPSGLDPEYQFTPSSSVIIGLDDKGEELLACDDKRIISVNA  
FT LIYGIS"  
FT join(complement(20071..>20265),complement(19912..19944)  
FT complement(19700..19829),complement(19540..19613))  
FT complement(19248..19459),complement(19096..19156))  
FT complement(18965..19018),complement(18800..18877)  
FT complement(18621..18726),complement(18372..18517)).
```

```

FT      complement(<18101..18211>)
FT      /gene="T8123.6"
FT      join(complement(20071..20263),complement(19912..19994),
FT      CDS      complement(19700..19829),complement(19540..19613),
FT      complement(19248..19459),complement(19096..19156),
FT      complement(18965..19018),complement(18800..18877),
FT      complement(18621..18726),complement(18372..18517),
FT      complement(18101..18211))
FT      /codon_start=1
FT      /db_xref="SPTREMBL:Q9FVU3"
FT      /gene="T8123.6"

Query Match      3.5% Score 133.2; DB 24; Length 90448;
Best Local Similarity 54.8%; Pred. No. 2.2e-10;
Matches 311; Conservative 0; Mismatches 248; Indels 9; Gaps 2;

QY 1407 accgatttttggcaagtatgtcgaattcttccatccacacacaaactaactaatta 1466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65632 AAGAAATATTCGCAAAAATTTAAATTAATTTCTATGACAAATTCGTGACGATCATTA 65573

QY 1467 gaactctagaagaataatttaaaatgaacaattatcgaaaaaattagttatgaaca 1526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65572 GCAGCTTATGCGCCGATGTATAGAAAATTTTGGCTAACTCAATTCCTAATATA 65513

QY 1527 attccagaattaacatga-----gcataataattaagaataaagaataatag 1579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65512 CTAGATTGATGCTCATGAAATCTTTGGGCTTTTCAGGAAATGATCAAGGCTTT 65453

QY 1580 atatgtatcgttaatacatatctgattcgcgttaaaaaaaacatatctgattaaattgctc 1639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65452 CTGGTGCMAAGTCTCTCGAGGCTGCCCTCCCTCAACTCGAGATGMAAGTITTA 65393

QY 1640 atgcagagcccatgcatcagatgcatcaacgcttttatttccacaataactatata 1699
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65392 ATTTCGGTATGAGAAATATTCCTTCAAGAAAGATGATGTGTAAGCTTTGAAACAGAT 65333

QY 1700 ttcaaaaaaagatttgcagattaaattttttgg--tgctgagcttcttccaact 1757
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65332 TGTTAGAAGTACGTCTCGATGGTGCGAGTCTCTTGAAAGATGAGCCACTTATCAATCT 65273

QY 1758 actaaactagtttggaaatgctctctcttcttatttcttcttcttcttcttctatgctt 1817
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65272 ACTAGATGATGTTAGAAATGCTTTGTGTGTTATTTTCGGGTTTCGCAATTCCTTATATTC 65213

QY 1818 ttacttaagaaattttaaagacgaattttagagcttctctcttcttcttcttcttctaa 1877
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65212 TCGTTATATATATGTAAGGATGTTAGTTCGCTCTCTCGTTACTCGCTTCGGA 65153

QY 1878 agtactcttggtaaacctcctcccaattagaacgctcaatcatatagttacttcta 1937
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65152 AATTTATTTTATAGACTATTTCTCTCCAAATTAAGCAATCATCATGAAATTAATATTTTA 65093

QY 1938 atatatgctagtcataaaaaaatga 1965
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65092 AATAATTTGCTGTTGATATAAAAAAAGA 65065

RESULT 9
AL139811 187576 bp DNA HTG 28-JUL-2001
LOCUS Homo sapiens chromosome X clone RP11-75A9 map pl1.23-11.4, ***
DEFINITION SEQUENCING IN PROGRESS ***, 3 unordered pieces.
ACCESSION AL139811
VERSION AL139811.23 GI:15021288
KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 187576)
AUTHORS Chapman,J.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

```

```

COMMENT
CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced g1:14586031.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA75A9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 2% of reads
Sequencing vector: plasmid; 108752; 97% of reads
Chemistry: Dye-terminator ET-amersham; 0% of reads
Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 187330 bases at least Q40
Consensus quality: 187366 bases at least Q30
Consensus quality: 187375 bases at least Q20
Insert size: 187376; sum-of-coverage
Insert size: 176827; 10.1% error; agarose-IP
Quality coverage: 7.90x in Q20 bases; sum-of-coverage
Coverage: 9.39x in Q20 bases; agarose-IP
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 93490: contig of 93490 bp in length
* * 93491 93590: gap of 100 bp
* * 93591 143600: contig of 50010 bp in length
* * 143601 143700: gap of 100 bp
* * 143701 187576: contig of 43876 bp in length.
* Location/Qualifiers
*   source
*     1..187576
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="X"
*     /map="P11.23-11.4"
*     /clone="RP11-75A9"
*     /clone_1kb="RP01-11.1"
*     /note="assembly-fragment:01965"
*   misc_feature
*     93591..143600
*     /note="assembly-fragment:03385..0"
*   misc_feature
*     143701..187576
*     /note="assembly-fragment:03385..1"
*
BASE COUNT 55497 a 38163 c 39269 g 54447 t 200 others
ORIGIN

Query Match      3.4% Score 129.8; DB 2; Length 187576;
Best Local Similarity 47.1%; Pred. No. 6.4e-10;
Matches 710; Conservative 0; Mismatches 767; Indels 30; Gaps

QY 1270 atgaattcttgtagcgaagaagtaaaatgaatgagcttcaacgcgaatatata 1329
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10974 ATACGTATATATATATATATATATATATATATATATATATATATATATATATATAT 11033

QY 1330 atttctgcanaactgaacttcaatcatgataatgataattcttcaattcccttg 1389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11034 AATTTAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 11093

QY 1390 ccaaatcaatcacatgcagagatttctgcgaagtatgctgctgcaatccctccatca 1446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11094 ATATATTTATATATATATATATATATATATATATATATATATATATATATATATAT 11153

QY 1447 -cgaacactaaacttaattagagctctaggaatatatlltaaaatgaacaattatcgaa 1505

```

[illegible]

	Db	12224	AAATATATATATATATATTTTATTCAGACACTATAATATAAAAATATATATATATATTATAGAT	12283
Cy	2571	agggaaaagcgaaa-attatctcaataatgacatatcccta-aadadaaaaaagaggc	2629	
Db	12284	ACTATATGTAATAAATATATATATATATATTTTAAATCTATATGTAAAAATATATATATAT	12343	
Cy	2630	atacgaugtaaatcataagaacatgcgatgttaagcattcggttaataatcagtcacaag	2689	
Db	12344	ATACTATATATAAAATATATATATATGATTTATAGATACATATATATATAAATATATATAT	12403	
Cy	2630	atatacatcacglaaacatcalatglatattcttggaaaagaatatatatattcggaag	2749	
Db	12404	ATATATCTACTATATATATAAATATATATATATATATATATATATATATATATATAAATAA	12463	
Cy	2750	taagaa 2756		
Db	12464	TAAATATA 12470		
RESULT 10				
AE001398/c				
LOCUS	AE001398	14867 bp	DNA	INV 06-NOV-1998
DEFINITION	Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.			
ACCESSION	AE001398			
VERSION	AE001398.1			
KEYWORDS	GI:3845197			
SOURCE				
ORGANISM	Plasmodium falciparum.			
REFERENCE	1 (bases 1 to 14867)			
AUTHORS	Gardner,M.J., Tetteh,I.H., Carnucci,D.J., Cummings,J.M., Aravind,L.Kooni,E.V., Shallow,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Singh,J., Aston,C., Lai,Z., Schwartz,D.C., Pethea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Hoffman,S.L. et al.			
TITLE	Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum			
JOURNAL	Science 282 (5391), 1126-1132 (1998)			
MEDLINE	99021743			
REMARK	Erratum:[PubMed] Erratum appears in Science 1998 Dec 4;282(5395) :[PubMed] (1827)]			
REFERENCE	2 (bases 1 to 14867)			
AUTHORS	Gardner,M.J.			
JOURNAL	Direct Submission			
TITLE	Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA			
FEATURES	Location/Qualifiers			
SOURCE	1..14867			
	/organism:"Plasmodium falciparum"			
	/db_xref:"taxon:5833"			
	/chromosome:"2"			
gene	ccplement(1570)..2424)			
	/gene:"PFB0490C"			
CDS	complement(1570)..2424)			
	/gene:"PFB0490C"			
	/note:"predicted by GlimmerM"			
	/codon_start=1			
	/product:"hypothetical protein"			
	/protein_id:"AAC71867.1"			
	/db_xref:"GI:3845198"			
	/translation:"MKENKIMDYVLSCPHDEVDRKKSGKNLSLKSSSTKKSIVY			
	SIISPKRSDSHHKGGSPGRGSFPINRKSGSEFKPYNNRLIKRNYYNKGKFNFIN			
	DNFGRTGSGSVFPDKRKSEFKRIISLNKNSVSYSYGNSKNKGRSKSFKAAPL			
	IIVTKRLNNTKITYASAPKRKNNINILSRKNRFALNTRKSKVCITKSSVPKRPR			
	GLKKSLSKSTRKSTSGSKPPKLNKYFLSKIKIIVTLNLKIPSLKEQKNTENVLP			
	NNATIRKN"			
gene	11241..14606			
	/gene:"PFB0495W"			
CDS	11241..14606			
	/gene:"PFB0495W"			
	/note:"predicted by GlimmerM"			
	/codon_start=1			

Db 549 AATCTATIAAAAAAATTAATTAATTTTATGAAATTAATAAATAAAGATTAATTA 608
QY 339 tgaagtcctcatcttgaaatttactcttcgtcttgctgatttcaatcctagaattga 458
Db 609 TTAATAAATCAATATATATATATTAATAATATTGATTATATTAATTAATTAACGATTA 668
QY 459 cgcacatatagaatttcattgctgacacagacatttagcgcgtatcctaattagacga 518
Db 669 TAAATATATTAATAATTTATTTTAATCATAATTCGAAATATATATATATATATAT 728
QY 519 ttaactagaataactatataataatagtggtgatttaagttaacttaaccactt 578
Db 729 AT 780
QY 579 gagacttcaacaatgggcaactacttaactgctgaaagaagcgctcccaactcgtgata 638
Db 781 AAATTTATTCGCCCTATTCATTAATTTATATATTAATTAACCTT-----AAAAAGTATT 836
QY 639 tgaagttatggggagagaccctgttaaactgcaataaacaacttaaggaactagaat 698
Db 837 TTTTAAAAAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 896
QY 699 tgaacccaaaatcataagaagacgttgaagaagatcattatcgtatccagctcatt 758
Db 897 TTTTATTAATTAATTTATTAATTAATTAATTAATTAATTAATTTTATATATTAATAATT 956
QY 759 tccctagcctaagatcaaatcaagcgcttgaagaagccttgaagaataatgacgaagaacc 818
Db 957 AA--AAATTAATTTTAAAAAATTCATTTTATATATATATATATATATATATATATAT 1014
QY 819 gtggggtttagaagaagaacagaagaatagaagaacatgctgtaaatgttcatttgg 878
Db 1015 TAAATTTCAATTAATTAATTAATTAATTAATTAATTAATTTTATATCCTAAATCTGA 1074
QY 879 tgaatagagagctgcaaaagagagagagagagaagaataggccaataatagagacact 938
Db 1075 AAT 1134
QY 939 aaaaatggagacatgctgttgatctactatataagaagcgactatgcttccctatgca 998
Db 1135 ATCTATAATTTATTTGATATATTAATAAATCTTAATAAATAATTTTAAAAAATAATGA 1194
QY 999 tgaatccaaacaaagtgcaacgctcctcttcttgccttaacttgcgaagctctctcc 1058
Db 1195 T-TTATTAATTAATTTATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTA 1253
QY 1059 tctctgcctctgaagaacaccctagaataatccctgctcctgactgctccttg 1118
Db 1234 TTTTAAATTAATTAATAAATAATTTTAAATATATTAATAAATGAATTTTAAATA 1313
QY 1119 agtaaccatgatttgcacacacactatctcttactcttcttggtcgtatagatlttg 1178
Db 1314 AAATTCATTCATATATTTATATATATATATATATATTAATTTTCAATTAATAATATAT 1373
QY 1179 cttaatactgcttctctgactgctcgtcgtacgtatacgaatttaagtgctataac 1238
Db 1374 AAGATATTAATAAATAATTTATTT-----TTAATCCTAAATCTGAATTAATTAATTTGATA 1428
QY 1239 aaggcttatataaactagcaaaatgctcagtaagaattgttagagaaagaagtgaaa 1298
Db 1429 TAT 1488
QY 1299 tataatgacttcaaacgcaatataataatttgcgtcaaaacttaacttaacat 1358
Db 1489 TAT 1548
QY 1359 gatcctaagatatttcttcaattctccttgcgaataatcaatcagcgaaggtcttg 1418
Db 1549 ATACTTAATTAATTAATTTTATTAATTAATTTTATTAATTAATAAT-----TATTTAAAA 1603
QY 1459 caagatagctgcaattctcctcctcacaacacacaaacttaatagaaccttagaa 1478
Db 1604 TAATTAATTAAGAAATATTTTATATATATATATATATTAATAATTAATTTTAAAAAATTCGA 1663

QY 1479 atatttcaaaatgcacacttatcgaaaaaaatttagtatagaaacaactccagaatt 1538
Db 1664 TTAATATTTAT 1723
QY 1539 aacatgagctataatcttaagaataaataatgaatlatatgataigtatgataacata 1598
Db 1724 TAAATAAATAATTTATTTTATATCCTAAATCTGAATTAATTAATTAATTAATTAATTA 1783
QY 1599 tctgattgcggtaaaaaaacaatcctgatttaattgttccatgcagcgcccatgcaacta 1658
Db 1784 TATATAAAAAATAAATAATTAATTTATTCGCCCTATTCATTAATTAATTTATATATATTA 1843
QY 1659 tgaatgcacacgcttcttattctcaca-----taactatataatcacaataaata 1711
Db 1844 ACTTAAAAAATTTTATTTTAAAAAATAATGTTTATTAATAATTAATTAATTAATTAAC 1903
QY 1712 gttctgcagatlaaattcttcttggtgcagcttctcccaacttaactaactagttg 1771
Db 1904 TTTTATATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1963
QY 1772 gaatgctcctcttattcttcttcttcttcttcttcttcttcttcttcttcttcttct 1831
Db 1964 AATATTAATAAATAATTTAATAATGATTTTATTAATAAATAATCAATTCATATATATAT 2020
QY 1832 ttaagacggaattggttagcttctccttcttcttcttcttcttcttcttcttcttctg 1891
Db 2021 TATATACATTAATTTATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTT 2080
QY 1892 acatcctccctcccaattagcagcagcagcagcagcagcagcagcagcagcagcagcag 1949
Db 2081 AATCCTTAATCTGAATTAATTAATTTGATATATATATATATATATATATATATATTA 2140
QY 1950 tgaataaaaaaatagaaaaaatactggcgtgactcactaaatgttggcuaaaaaatc 2009
Db 2141 AAATTAATAATTTATTCGCCCTATTCATTAATTAATTAATTAATTAATTAATTAATTA 2200
QY 2010 gatataagaaatcccaatcacttcttcttcttcttcttcttcttcttcttcttcttct 2069
Db 2201 TTTTAAAAAATAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTTTAA 2259
QY 2070 ttaataaataaagacttgaatgttgaatgttgaatgttgaatgttgaatgttgaatgtt 2129
Db 2260 TTAATTTTAATAATTAATAATTAATTAATAATTAATAATTAATAATTAATAATTAATA 2319
QY 2130 aactatcctctggaataa--ctgaagaatttaactatcttcaaaatlaaacttagctc 2187
Db 2320 ATATATTAATAATTAATAATTAATTTAAATTTAAATAATTAATAATTAATAATTAATA 2379
QY 2188 cgaataatacagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2247
Db 2380 TTAATTAATAATCTATTCATTTATATATTTAAATTAATAATAATAATAATAATAATA 2439
QY 2248 tgaatacaaacctataaataactaaatgctcgtcgtggttcaaaagccttccaag 2307
Db 2440 AATTAATAATAATTTATTTATTAATTAATTAATTAATTAATAATAATAATAATAATA 2499
QY 2308 tctcatcttctt---aaatcgaagacatttcaatagaagaatagctgagcagcaaaaaa 2364
Db 2500 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2559
QY 2365 taatggtatcttgaaggttcttcttcttcttcttcttcttcttcttcttcttcttct 2422
Db 2620 TTTTAAAAAATTTCTTAATATGATATATTTTATTAATAATAATTTATTAATAATAAT 2679
QY 2483 taataagaatataaataacaaatataataataacttaaaatcttaacataaactc 2542
Db 2680 TTTTAAAAAATAAACAATAATTTTAAATAATTAATTAATTAATTAATTAATTAATTA 2739

[illegible]

```

REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Parquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of
JOURNAL Drosophila melanogaster mitochondrial DNA
MEDLINE Mol. Biol. Evol. 11 (3), 523-538 (1994)
94285822
REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the
JOURNAL nucleotide sequence and evolutionary comparisons
MEDLINE Insect Mol. Biol. 4 (4), 263-278 (1995)
96423163
REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Departmen
Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
source
1..19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/db_xref="taxon:7227"
/note="derived from new and previously submitted
sequences; sequence is a composite containing sequences
obtained from different Drosophila melanogaster strains
1..65
/product="trna-ile"
/complete(97..165)
/product="trna-gln"
171..239
/product="trna-phe"
240..1265
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MENNSSKILFIITMIIGTLITVTSNMLGAMGDLINLSF
LSDNNNASTREASAKYELTQVLASTVLEFSEFSLMLKNNNNEISFSTIIMS
LSCAPAFHFPMPNPMAGCTTVMANLMTMOKIAPMLISXIKYILIVSLIS
GATGGLNQTSLRKLMATSSINHLGMLSLMISESTILILFFYSLSPLYTTFMF
KLHLNOLFSWFFVNSKILKFTLTFMNEUSLGLPPGLGFPKWLIVDITLGNVFF
IMMSTLITLFFLRICYSAFNMNFEENNAIMKMNNSINYNMIMTFPSIFGL
SLEFMP"
1264..1329
/product="trna-trp"
/complete(1322..1383)
/product="trna-cys"
/product="trna-tyr"
join(1470..1472,1474..3009)
/note="mechanism underlying reading frame shift after
first codon uncertain"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"
/db_xref="GI:7412849"
/translation="MSRWLFSTNHRDICTLYFEGAMGMYGTSLSIIIRAEIGL
ALIGDDIYVITAFNAFLIMFFNVRIMIGGNNILVPLMLCAPOMAEPRNNMMS
LIPDPAISILVSSMENGASTGWYVPIASGIAHGASVDLAFSLIILAGISII
VNIITIVNMRSGISLDRNPIFVMSVITALLLSLPLACATMMLDRNLN
FPDAGGDLPLVGHILFMFGHPREYVILIPGFEMISHIIISQSGKETTESLQAT
LAIGLGLFIYMAHMFITVCGDVDTIRAFVTSALMIAVPTSIKIFSMALATLHGTL
PALMALGVEFLFTVGGILGVVLANSVDILHDITYVVAHFVYLSMGAVFIMIM
IHMYFLPTGILNNKMKLSHFIIMELGVNLTFRPOHFLGAGRPARYSDPAYT
IVSTIGTISILCLFFHFIIMESLVSQRQVVPIDNSIEIYQNTNPPAFHSYS
LTN"
3012..3017
/product="trna-leu"
3083..3767
/note="trna stop codon is completed by the addition of 3
residues to the trna"

```

```
/codon_start=1
/transl_except=(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTMANLGIQDSASPLMEQLFFHDAALLIVMTIVLVGLMFM
LEFNVYNRELHGLQLEIMWTILPAIILFLALPSLRLLYLDLEINPSYLSKSIQH
OMWMSYSDFNINIEDSYMIPTNELMDGFLLDVDNIVVLPNNSQIRILVTADVI
HSMTVPALGVKVGCTGRLNQTNEFINRGLFGQCEICGANHSFMPIVIESVNVY
FIKWISSNNS"
3768..3838
/product="tRNA-Lys"
3840..3906
/product="tRNA-Asp"
3907..4068
/codon_start=1
/transl_table=5
/product="Atpase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQKAPISMLLFLIFISITFLFCSINYSYMSPKSNELKNI
NLNSMKMK"
4062..4736
/codon_start=1
/transl_table=5
/product="Atpase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="MNTNLFVSVDPLAIFNPSLIMLSPFICILMIPSYIMLPSYNI
MNSILITLHKPKLIGSGHNSSTFIFISFSLTNNPGLPEPYIFTSNHLITL
LSLALPLMLCPLMIGXINHTOMHFAHLYOGCPAILMPVVCITETISNTRGTLAVR
LRNMTAGHLILLGNTGSSMSMLMTFLMLAQIALLVESAVALIOSYFAVESTL
YSEVN"
4736..5524
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/translation="MSTSHNHPFHLVDYSPMLTGAIGAMTVSGMVMFHOYDLSLF
VLNITLITLVQMMDVREGTYOGLHTAVTIGLRMGMLILSEVLFEFSWAF
FHSLSLPALEIGAMPMPGIIISPNPOIPLNTALILASGVTVTAHSHMNNNSOT
TQGLFTVLGIYFTILOAYEYIEAPETIADSIYSTPMATGFGHIVLIGITFLV
CLIRHLNHFSSKNNHGFPAALAMVMHFDVVMFLYITITVMGG"
5543..5607
/product="tRNA-Gly"
5608..5961
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/protein_id="AAC47817.1"
/db_xref="GI:1166536"
/translation="MFSIIFIALLLILLITIVMFLASILSKKALIDREKSSPFECGF
PKSSSLRPLSLRFFLITIFLIFDVEIALILPIMIMKYSNIMWTITSIIFITILLI
GLYHEMNOQMLMNSN"
5981..6045
/product="tRNA-Ala"
6055..6118
/product="tRNA-Arg"
6119..6183
/product="tRNA-Asn"
6184..6251
/product="tRNA-Ser"
6252..6318
/product="tRNA-Glu"
complement(6337..6401)
/product="tRNA-Phe"
complement(6401..8124)
/Note="tRNA stop codon is completed by the addition of 3' A
residue to the mRNA"
/codon_start=1
/transl_except=(pos:complement(6401..6402),aa:TERM)
```

```
Query Match 3.28; Score 125.2; DB 3; Length 19517;
Best Local Similarity 44.48; Pred. No. 3,5e-09;
Matches 1187; Conservative 0; Mismatches 1458; Indels 49; Gaps 1

QY 99 ttggaaccgaacatcctaatttggaaagaaaaaccctatgtaataatcacctgaagc 158
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15228 TTTTITTTTAAAAAATTTTITTTTAAAAAATTTTITTTTAAAAAATTTTAAA 15287

QY 159 aaaaaaatatcctcatcatataatttcttaagattatcatcttcctgtgtaa 218
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15288 AATATATAGATTAAATTTCTTTTAAATGACTAAAAAAATTTTITTTTAACTATTAAA 15347

QY 219 gaggctaccatctgcaagctgtgtatctaataaaagagcaagtaagtaagcctta 278
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15348 ACTITTTT---TTTACAAATTTTAAAAAATATATATATATATATATATATATAT 15404

QY 279 taataatlgatgttcagtagtaactggtatctgttgatgaataaacaagatgaacat 338
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15405 TTTTAAAAAATAAATGAAATTTATATATATATATATATATATATATATATATAT 15464

QY 339 taatgcatgaagcgagcttctgtgtaaccttaagaagcaatgtgaagaagatg 398
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15465 AATCTATATAAAAAATTTATTAATTTTATGATATATATATATATATATATATATAT 15524

QY 399 tgaagctccatttggaaaattatcttc:agcttgcgactt:taaatctagaatgaa 458
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15525 TTTAAAAATCAATATATATATATATATATATATATATATATATATATATATAT 15584

QY 459 cggacaataagaaatctcagctgagcaaacgacttaagaagatccttaattgaacga 518
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15585 TAAATATATATATATATATATATATATATATATATATATATATATATATATAT 15644

QY 519 ttaattgaataatactatataataatgtaagtgatgaattgaattactaccct 578
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15645 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 15696

QY 579 gaggaaatt:aaacaatgggcaata:acctaaagtcgaagaagcgctccaccct:cgtgtaa 638
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15697 AAATTTATCCCCATTCATTAATTTATATATATATATATATATATATATATATAT 15752

QY 639 tgaagttagagagagagctcc:gttaactgcgaataaacaacttaagaagcagaat 698
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15753 TTTTAAAAAATAAATTTATATATATATATATATATATATATATATATATATAT 15812

QY 699 tgaacacaaatacataagaagaaagcttgaagaagctcatatcgtatccagctcatat 758
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15813 TTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 15872

QY 759 tccctagctaaagatcaatacgaagcgcttgaagaagcctgt:aaagaaa:gtcgaagaaac 818
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15873 AA--AAATTAATTTTAAAAAATTCATATATATATATATATATATATATATATAT 15930

QY 819 gtagagcttagaagaagaagaataagaacaacatgaagcttaacttgcctttag 878
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15931 TAATTTTCAATTAATTTATATATATATATATATATATATATATATATATATATAT 15990

QY 879 tctatagaagcttgcgaagaagagagaggaagaagaatacctagtcgaataatgaagact 938
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15991 AATATATATATATATATATATATATATATATATATATATATATATATATATATAT 16050

QY 939 aaaaaatgagagacatgctgttgaagtaactatacagaagcagacttgcctttagc 998
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16051 ATCTAATAATTTATTTGTAATTAATAACTTAAAAAATATTTTITTTTAAAAAATAATGA 16110

QY 999 tgaataccaaccaaagtcgaagcctctt:ttgcctcaatttcgtaagctctcctcc 1058
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16111 T-TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 16169

QY 1059 ttcttcgctcttaggaagaaccccaagaatctaactccctgtcttcgtccttcttg 1118
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16170 TTTTAAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 16229
```


protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
http://combio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-se
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K8A10 and the 3' clone is K3G17.

FEATURES

source

1. 74342

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="K21H1"

/complement(join(683..778,1044..1247,1325..1959,2053..2283,

2364..2661))

/note="gene_id:K21H1.1

unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB10939.1"

/db_xref="gi:10177593"

/translation="MANSDDDERGCVILRPPELRPMELVRLPFGSDIKPRSDSS

ETEHSEFRHMLIFVSLVLKLFPSKLLALVGSALFSLNLSNNSFGLRGV

VMPQTSNYSQVIFGLHDLTVSLDMITNREDEGYALSLMASKIAYENARLKHV

ENHNMKMYLGLVDYNEVYOEKETOAFIMSDDETTTVAAPFGTELPNE

DMCSDFDTFPELPLNIGNIGGFMKALGONNCSMPKPSNPDKSPLAAYSLRDL

KTLIAONKNTKFNVLGHSIGALALIFPAVVIHHTPELLERIGVYVTCGPPRCDK

FGFERMEKLEKYNIKYRFVYNNDIYVRLPIDDDMLFRRHGTCTIYDQNYAKVRE

QSDENFLRLGRIKMMMSAILEFIRSTYVAEKSEYSEGLKLGKRALGIIYVGVSN

HTPDYKNSRVASINGEDYPRKDETLAKRVIFIPG"

7772..8497

/note="emb|CAB62312.1

gene_id:K21H1.2

similar to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB10940.1"

/db_xref="gi:10177593"

/translation="MDSIDIMMMHMHOMKLEPFCNPNSFFSDHNTTYPLFNSTHY

OSDSMTNEPGRKSGSLTNPSSISPTAYSSVLDKRNNNNNNNTNMAAREHI

FRIVAMOPIDHDEPAVKPRKRRNRYISKDPSVAARRRREIRIRIILRIYVCGTK

MDTASMLDEAIHYVKFLKKVOSLEQAVYVGGGGGGGGRVLIIGGSGMTAASGGGSGG

GVNKKCGTCTVTHQVGNAILR"

9616..10005

/note="gb|AA64534.1

gene_id:K21H1.3

similar to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB10941.1"

/db_xref="gi:10177594"

/translation="MAASLNLILLISLTLFFISLQSESLSDNSLTLPLDGFEDWP:3

HSDEFDIDGESFEVTEEDGVTRDSILYRRRTKYISYGALSANRVPCCPPRSGRSY

YTHNCFRARGVHPYRSCSSITRCRR"

14222..15256

/note="gene_id:K21H1.4"

/codon_start=1

/evidence=not_experimental

/product="protein kinase-like protein"

/protein_id="BAB10942.1"

/db_xref="gi:10177595"

/translation="MEMIRGETIGYGTFTSYSLATRSNNDGFEPLMAVKSADSYA

ASLANEKSVLNLGDCCNEIYRCGEORTYENGEMHNLLEYASRGSLSYLRKLAG

CDS

complement(15607..17817)

/note="gene_id:K21H1.5"

/codon_start=1

/evidence=not_experimental

/product="subtilisin-type protease-like"

/protein_id="BAB10943.1"

/db_xref="gi:10177596"

/translation="MGMVTVLIVSFVFAIVTAETSPYIIMDLSANLPFSDHRS

FSTTLVSTNNKPKIIVATDSVGSALVILNSELQRLKHKPGVSEFTKLPYLAIH

TSPKFIQVNSTGMPVNSAGAVIGIIVDIPDSFDDDGVSFSPKWKAC

FNSSSLCNKILGAVENKGLFANNPDLRETIGQYSPYDIGHGIVAAIACNH

KNASFSYACGTASGIAFPAHLAIYKAAMEEITSSDYVIAIDQAIROGVHISL

LSFEDDDNDGFGLENDPIAVASFALIOGVFVTSGGNDGYVSLNGAFVIMTV

AGTIGRQFGTILFGNRSEFSFSPGEPVQRPVYIEGSEVNTFLANRIVC

ENINIGSKLHQIRSTGAAAVVLIIDKLEEDPTIKFQPVAFVIGSKHETTESVAS

KNNAKLEFRRTVIGTKAPAPGVYSSRGPTIPQILKPDILAGTLLSAMPV

OITGTRALPLFSGPNLTIGTSMADPVAAGVALLIOVHPNMSPAIKSAIMTALT

NPLVAGCHVSTNKNVLPGLIYDTPRODINFLCHFAKOSKRLNIITRVSQAK

PEPYLNPSTIAVFTSDOSPRTIFRITLNNGEAKRSYIVRGKLTNNVPEPKI

FSEKNKLSYTRLESFRLGLENVYGLVSWDEDEHNSCVYVTSVLES"

join(18490..19596,19689..19859,19945..20028,20117..20

20278..20342,20515..20605,20698..20883,21116..21279,

21388..21502,21780..22146,22391..22510,

22590..22904,22985..23074,23461..23579,23668..23740,

23831..23887,23999..24091,24221..24343,24422..24519,

24598..24633,24730..24843,24949..25032,25116..25322,

25414..25490,25698..25821,25946..25999,26108..26227,

26298..26372,26460..26600)

/note="gene_id:K21H1.6"

/codon_start=1

/evidence=not_experimental

/product="DNA polymerase alpha 1"

/protein_id="BAB10944.1"

/db_xref="gi:10177597"

/translation="MSGQNTETGRRRSGAFASSSRKDTLERLKAIRGGIRASGG

YDRLQKPIFDVDEEDYDALSTREREARGVVEDEGGDGIYDEGEELERKPS

PSTDESDGGRSGRLKKKKKQDOTQOYKKNPALKAATITGEGSLISMTS

SPKVKETDKAOYEGILDEIIVQVTPDESDEKRRRLKPGVTPYIKRKLFSVA

SMOKSEFPPTTYEGDSYSHNDELKEDMESVYPSIELIGSDIYKDGSKN

KRTVKSLEIGVFEFTLNKATIDMKEDSLSTAGKREMGVGTENGLJSSSRG

TEFDLADGSLERFIIDAYEAFGASMGTYIFGVKMDIYKSCVAVKNIQKCY

IENDSIFPSHELIIMEQEVKDSRLSPESFRGLHMAKLNKIQE:LIQVSNFS

APVKRNYAFEDRPVAGEYVLYKINSFEDRLPEDLKESFALGSHTSLEHTI

KKRIQPCMLKISSFTGSPSEGVAMCKFEVYVQPKDITIVSEKVVHPAVVTA

NKTIVNQRQNTSEFVSASVCEHNAKIDVHPAPARSRSGSLISHTVVRNPEGIGY

IQMKKEVSQRNKNQCNVYSFNSRRALLNRLFLINKLIDSDIILVGHNTSCFDLVL

QAOAKCKVSSSMWKSITGRKRSFPRKLGNSNYSGATPGLMSCIAGRLCDPTDLS

DILKEVYSVLDLSLTQ:NRDRKELTAPMDIPMFOSSTVLELIECGFTDMLSEL

FHLVPLTLQ:LTNLSGNLMCKTLOGAQAQRIEYLLHTFHSKRFILDPKDISQAKK

KSKRRMDYAPEDRNVDLADLTLENDPSKSKTKKPAAGVLYLSPKRLVDKY

LILDEFSIYPIIIOENICFTIIPRSDGVPLPSSQIPGLIKLMEHLISIRKSV

KKKKTGLKYMELDIRQALKLTANSWKGCGFSNRYAKRLAELITLGGDITG

YDLYONHNLNLYVIGTDSIMHSGIDILEYKAIKSVIIOENKRYKCIKIDCGI

KMLRLRKKRYAAYVLOERDKSPCEDIERKGVDRPMSLSKTIQDLSKTIYIG

SCEDVEALIHNEMLIKERNRGVALKRYVITKTLTPPAPYPSKQSPHVOVLR

RORGYKKEFNARDIVPIIICYEQAGNASSASGIAERARHDEVSSESRMLVDIDY

LAQOIHVVSRLCAEIOGTSFERLAECGLDPSKRSNATSSDPSLSLFATSD

ESKRPATPEESDFTWKLICPGQOODEGIGISPAMIANOVKRODGEVSMYK

IIVCEDSCSKHTTBSNPFILGERRGVCPVPCNGTLLRKYEADLYKQLSYVC

IIDQCGLSEKMGVRIYVERKAMTKIRAPVYAAAITSSRRRCAYGMQGLDVIY1"

complement(join(26828..27019,27114..27179,27258..27323

27414..27485,27598..27834))

/note="emb|CAB1798.1

gene_id:K21H1.7

similar to unknown protein"

CDS

complement(15607..17817)

/note="gene_id:K21H1.5"

/codon_start=1

/evidence=not_experimental

/product="protein kinase-like protein"

/protein_id="BAB10945.1"

/db_xref="gi:10177595"

/translation="MEMIRGETIGYGTFTSYSLATRSNNDGFEPLMAVKSADSYA

ASLANEKSVLNLGDCCNEIYRCGEORTYENGEMHNLLEYASRGSLSYLRKLAG

```
/db_xref="GI:10177598"
/translation="MGDSVVGRLPPSSDELSSFLROIILSRPTAPSPSPKSTNV
SSAEFFPSVSGAVSVGVSEGTGODKAFPHKRSKAKNSLRINDAOPNLSE
KRRSRINKMKAKLILPNSNKTDKASMLDEALIEIKLOLOVOT.LAMNGLINPM
RLPOVPPHTIRINLETBODLNLFTLLAHPHLEPAKTSOGCFSFATLL"
complement(join(29326..29586,29960..30082,30390..30512,
30845..30958,31036..31233))
/note="gene_id:K2IH1.8
pir||T05432
similar to unknown protein"
/codon_start=1
/evidence-not-experimental
/protein_id="BAB10946.1"
/db_xref="GI:10177599"
/translation="MDPNEDNSEIPDPSPFMWLLFGDDMDLMDTSPYDAUDISPP
TLVDGDAIPRGSILRTISMETFEODLSQMDPOSETTSVVEETDVSQKNTY
NEETLVRWYRWLQSSSTNSETSGSEPTMASSHGFIYEETLEDRHNLQSS
7NEOFTEBLETETNPSSHGDFTTEBELSTIDETGDEGTGSEEVIDENTLRRYEK
RSDDETRKRVICQALKDNEVSKLGGCGDFHFGCIKMLMTNCPCLCNREV"
join(33235..33402,33654..33791,33873..33971,34064..34367,
34837..35225)
/note="contains similarity to MAP3K-like protein kinase
gene_id:K2IH1.9"

Query Match      3.2%; Score 125.2; DB 8; Length 74342;
Best Local Similarity 64.5%; Pred. No. 3.2e-09;
Matches 187; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 3015 acatgatgatgaagaatcagtlactatgctgcgcgcacgcagcccgatgaatcg 3074
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8052 ACATGCAGCTATGCGAGKATATATCTTCGTAATCGCCGATGCAACGCATCATATCG 8111
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3075 accctgcacagctccctaaaccgaaccgcctaaacgtaagaacgaacatcccca 3134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8112 ATCCCGAGGCGGTTAAGCCACGACAGAGAGAAAGTCTCTAAGATCCTCAAA 8171
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3135 cagtggtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8172 GCCTGGCGGCTAGGCAATAGAAAGGAGAGAAATAGCGAGATTCGATTTTCAACGCG 8231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3195 tcgtgcctggtgctgaagaatgacacagctccatgctgcgcgcgcgcgcgcgcgc 3254
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8232 TTGTTCTGCTGGACGAGATGATGATACGCTTCGATGCTCGATGAGCAATCATATTATG 8291
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3255 ccaagtcctgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8292 TGAAGTTTAAAGAAACAGCTGCTCTGTGAGAGACGACGCGGTGTT 8341
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
AC092304/c 36977 bp DNA HTG 03-JUL-2001
LOCUS Homo sapiens chromosome 19 clone LNLPOS-20C5, WORKING DRAFT
DEFINITION
AC092304
AC092304.1 GI:14589493
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 36977)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 36977)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
```

```
web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 31883_FOS36853
Center clone name: LNLPOS-20C5
-----
Summary Statistics
Consensus quality: 3111 bases at least Q40
Consensus quality: 33034 bases at least Q30
Consensus quality: 33871 bases at least Q20
Estimated insert size: 42960; agarose-1p estimation
Estimated insert size: 36477; sum-of-ctdigs estimation
Quality coverage: 6.56 in Q20 bases; agarose-1p estimation
Quality coverage: 7.73 in Q20 bases; sum-of-ctdigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1580: contig of 1580 bp in length
* 1581 1680: gap of unknown length
* 1581 3748: contig of 2068 bp in length
* 3749 3848: gap of unknown length
* 3849 6070: contig of 2222 bp in length
* 6071 6170: gap of unknown length
* 6171 13635: contig of 7465 bp in length
* 13636 13735: gap of unknown length
* 13736 22951: contig of 9216 bp in length
* 22952 23051: gap of unknown length
* 23052 36977: contig of 13926 bp in length.

FEATURES
Source
1..36977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LNLPOS-20C5"

BASE COUNT 1:122 a 7407 c 7254 g 10688 t 506 others
ORIGIN

Query Match      3.2%; Score 124.8; DB 2; Length 36977;
Best Local Similarity 45.3%; Pred. No. 3.8e-09;
Matches 745; Conservative 0; Mismatches 873; Indels 27; Gaps

QY 1165 tctataggaatttgccttatatgcttcttgatgctccgaagtaacgaatt 1224
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34403 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 34344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1225 caaatgctatatacaagcttatataacacacacacacacacacacacacacacacac 1284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34343 AATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 34284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1285 gaaaaagtagaataatataatgcttctttaaagcgaacataataatcttgcgcaaac 1344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34283 GTATATATACATATATATATATATATATATATATATATATATATATATATATATATAT 34224
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1345 ttacgttaacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 1404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34223 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 34164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1405 gaacggaattctggcaagcttaagctgaacatctccatccacacacacacacacacac 1464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34163 ATGTTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 34104
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1465 tagaactctagaacaattttaaacatgaacattcttgaacaaatllagcttgaacaa 1524
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34103 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 34044
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1525 caattccgaatlaaacatgagctatataatlaagataaactgaagtaacatgatactg 1584
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Mon Apr 1 08:29:48 2002

us-09-548-971b-2.rpt

A:Cross-references: GB:AE002093; NID:q4115386; PIDN:AA03387.1; GSPDB:GN00139
A:Gene: At2g24260
A:Map position: 2

Query Match 14.9%; Score 157.5; DB 2; Length 350;
Best Local Similarity 44.2%; Pred. No. 4.5e-06;
Matches 34; Conservative 16; Mismatches 22; Indels 5; Gaps 2;

QY 109 PKRRNRVRI-----SDPQTVARRRRERISEKIRILKRIYVPGAKKDTASMLDEAIRYT 64
DB 131 PPGSRTRIRARRGQATDPHSIAERLRERIERAKALQELVPPGNKTKDKASMLDEIDYV 190
QY 165 KFLKROYALLOPHSQIG 181
DB 191 KFLQVYKVLMSRRLC 206

RESULT 10
T01090
hypothetical protein T10P11.13 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999

C:Accession: T01090
R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffman, J.; Till, S.; de la Bastide, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.
Submitted to the EMBL Data Library, November 1998
Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.
Accession number: 214248
Accession: T01090
Note: Translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Molecule: 1-310 <KAP>
A:Residues: 305; EMBL:AC002330; NID:q2262135; PID:q3892050
A:Cross-references: cultivar Columbia
A:Experimental source:
A:Genetics:
A:Map position: 4 4/3; 206/3; 233/3
A:Mutations: 121/3; 162/3; 18;
A:Note: T10P11.13

Query Match 14.8%; Score 156.5; DB 2; Length 310;
Best Local Similarity 33.6%; Pred. No. 4.7e-06;
Matches 39; Conservative 21; Mismatches 18; Indels 9; Gaps 2;

QY 78 DEYEDMDAMKEMQYMTAVMQPVDIDPATVPR 131
DB 109 DDVYNRCSSMKRPFVHGPRMQP---PPSAPRNPRT 165
QY 132 ERISKIRILKRIYVPGAKKDTASMLDEAIRYTKFLKROYK 187
DB 166 ERIRERIRALZELVPTVTKTIDRAMIDELVDYVKFLRLQYKVL 221

RESULT 11

hypothetical protein AA025805.1 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86161
R:Thielmann, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, V.; T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizier, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, A.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: B86141; M01D:21016719
A:Accession: B86161

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE005172; NID:q4587574; PIDN:AA025805.1; GSPDB:GN00139
A:Map position: 1

Query Match 14.1%; Score 149; DB 2; Length 297;
Best Local Similarity 39.4%; Pred. No. 2e-05;
Matches 39; Conservative 15; Mismatches 35; Indels 10; Gaps 2;

QY 98 MQPVDIDPATVPRPN-----RRNRVRI-----SDPQTVARRRRERISEKIRILKRIYVPG 107
DB 117 MKPPMSQAPAP-PMPHOOSTIRPVARRGQATDPHSIAERLRERIERAKALQELVPPGNKTKDKASMLDEIDYV 190
QY 149 AKMDTASMLDEAIRYTKFLKROYRILLOPHSQIGAPMANP 197
DB 176 NKTDRAAMIDELVDYVKFLRLQYKVLMSRRLCGAGAVAP 214

RESULT 12

hypothetical protein (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86228
R:Thielmann, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; W. Chung, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, V.; Jensen, N.F.; Hughes, B.; Huizier, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: B86141; M01D:21016719
A:Accession: H86228
A:Status: preliminary
A:Genetics:
A:Molecule type: DNA
A:Residues: 1-524 <STO>

A:Cross-references: GB:AE005172; NID:q3482928; PIDN:AA033213.1; GSPDB:GN00139
A:Map position: 1

Query Match 13.6%; Score 143.5; DB 2; Length 524;
Best Local Similarity 28.8%; Pred. No. 0.00011;
Matches 53; Conservative 22; Mismatches 36; Indels 53; Gaps 2;

QY 7 KKGVCSCVSSKRSKSNHSPKSMERPOPHLLMDNKKANDLTTQHAVALNDPNNHMLMD 277
DB 277 EKAVVSCSVSCNSLSDPSPSLSKRKSNINQD-----IDCH----- 524
QY 67 PPETLILHDEDEYEDMDAMKEMQYMTAVMQPVDIDPATVPRPKRRNRVRISSDPQTV 316
DB 316 -----SEDE-EESSGGRKE-----AGSRTGLGSKRS--RSAEVHNLS 524
QY 127 ARRRERISEKIRILKRIYVPGAKKDTASMLDEAIRYTKFLKROYRILLOPHSQIGAPMAN 352
DB 352 ERRRRDINKKALQELVPPGNKTKDKASMLDEIRYTKFLKSLQVQVQVMS-----MAS 524
QY 187 PSYL 190
DB 404 GYVL 407

RESULT 13
B86432
Protein f518.12 (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001